ICSNC 2013

The Eighth International Conference on Systems and Networks Communications

ISBN: 978-1-61208-305-6

October 27 - November 1, 2013

Venice, Italy

ICSNC 2013 Editors

Renzo Davoli, University of Bologna, Italy

Josef Noll, University of Oslo & Movation, Norway

ICSNC 2013

Forward

The Eighth International Conference on Systems and Networks Communications (ICSNC 2013), held on October 27 - November 1, 2013 - Venice, Italy, continued a series of events covering a broad spectrum of systems and networks related topics.

As a multi-track event, ICSNC 2013 served as a forum for researchers from the academia and the industry, professionals, standard developers, policy makers and practitioners to exchange ideas. The conference covered fundamentals on wireless, high-speed, mobile and Ad hoc networks, security, policy based systems and education systems. Topics targeted design, implementation, testing, use cases, tools, and lessons learnt for such networks and systems

The conference had the following tracks:

- WINET: Wireless networks
- HSNET: High speed networks
- SENET: Sensor networks
- MHNET: Mobile and Ad hoc networks
- AP2PS: Advances in P2P Systems
- MESH: Advances in Mesh Networks
- VENET: Vehicular networks
- RFID: Radio-frequency identification systems
- SESYS: Security systems
- MCSYS: Multimedia communications systems
- POSYS: Policy-based systems
- PESYS: Pervasive education system

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

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

This event could also not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the ICSNC 2013 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success. We gratefully appreciate to the technical program committee co-chairs that contributed to identify the appropriate groups to submit contributions.

We hope the ICSNC 2013 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in networking and systems communications research. We also hope the attendees enjoyed the charm of Venice.

ICSNC 2013 Chairs

ICSNC Advisory Chairs

Eugen Borcoci, University Politehnica of Bucarest, Romania
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Reijo Savola, VTT, Finland
Leon Reznik, Rochester Institute of Technology, USA
Masashi Sugano, Osaka Prefecture University, Japan
Zoubir Mammeri, IRIT, France

ICSNC 2013 Research Institute Liaison Chairs

Song Lin, Yahoo! Labs / Yahoo Inc. - Sunnyvale, USA
Habtamu Abie, Norwegian Computing Center - Oslo, Norway

ICSNC 2013 Industry/Research Chairs

Rolf Oppliger, eSECURITY Technologies - Guemligen, Switzerland
Jeffrey Abell, General Motors Corporation, USA
Christopher Nguyen, Intel Corp., USA
Javier Ibanez-Guzman, RENAULT S.A.S. / Technocentre RENAULT - Guyancourt, France

ICSNC 2013 Special Area Chairs

Mobility / vehicular
Maode Ma, Nanyang Technology University, Singapore

Pervasive education
Maiga Chang, Athabasca University, Canada

ICSNC 2013

Committee

ICSNC Advisory Chairs

Eugen Borcoci, University Politehnica of Bucarest, Romania
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Reijo Savola, VTT, Finland
Leon Reznik, Rochester Institute of Technology, USA
Masashi Sugano, Osaka Prefecture University, Japan
Zoubir Mammeri, IRIT, France

ICSNC 2013 Research Institute Liaison Chairs

Song Lin, Yahoo! Labs / Yahoo Inc. - Sunnyvale, USA
Habtamu Abie, Norwegian Computing Center - Oslo, Norway

ICSNC 2013 Industry/Research Chairs

Rolf Oppliger, eSECURITY Technologies - Guemligen, Switzerland
Jeffrey Abell, General Motors Corporation, USA
Christopher Nguyen, Intel Corp., USA
Javier Ibanez-Guzman, RENAULT S.A.S. / Technocentre RENAULT - Guyancourt, France

ICSNC 2013 Special Area Chairs

Mobility / vehicular
Maode Ma, Nanyang Technology University, Singapore

Pervasive education
Maiga Chang, Athabasca University, Canada

ICSNC 2013 Technical Program Committee

Habtamu Abie, Norwegian Computing Center - Oslo, Norway
Joao Afonso, University of Lisbon, Portugal
Jose Maria Alcaraz Calero, University of Valencia, Spain
Pedro Alexandre S. Gonçalves, Escola Superior de Tecnologia e Gestão de Águeda, Lisbon
Abdul Alim, Lancaster University, UK
Sultan Aljahdali, Taif University, Saudi Arabia
Abdullahi Arabo, Oxford Internet Institute / University of Oxford, UK

Copyright Information

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

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

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

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

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

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

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

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

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

Table of Contents

A Design of Full-Rate Distributed Space-Time-Frequency Codes with Randomized Cyclic Delay Diversity

Hong-yu Fang, Meng-lin Bao, Lei Xu, Xiao-hui Li
Key Lab of Intelligent Computing and Signal Processing of Ministry of Education
Anhui University
Hefei, China
e-mail: xhli@ahu.edu.cn

***Abstract*—Cyclic Delay Diversity (CDD) method was introduced to cooperative communications for improving the system diversity performance. Since the CDD with fixed cycle delay cannot obtain the optimal system performance in all situations, this paper presents a full-rate distributed space-time-frequency codes scheme in the full-rate cooperative communication model by taking advantages of Randomized Cyclic Delay Diversity (RCDD) method and Linear Constellation Precoding (LCP) technology. The proposed scheme is more practical and has the advantage of low detection complexity. Compared with the full-rate cooperative communication scheme with Fixed Cyclic Delay Diversity (FCDD), the proposed scheme can achieve Better Bit Error Rate (BER) performance in the case of large number of subcarriers.**

Keywords-randomized cyclic delay diversity; cooperative communication; OFDM; space-time-frequency code

I. INTRODUCTION

The traditional two-hop cooperative mode requires two time slots to complete data transmission. It increases the system diversity gain at the cost of half of the system transmission rate [1][2]. In order to solve the problem, the Non-orthogonal Amplify and Forward (NAF) transmission mode with single relay was proposed by Nabar et al. [3]. With the help of the relay node, the source node transmits two symbols to the destination node within two adjacent time slots. Thus it can achieve full-rate data transmission. However, only the data sent from the source node in the odd time slot is forwarded at the relay node results in the imbalance of error rate between odd and even time slots, which is also known as "short-board effect". In order to overcome this phenomenon, linear constellation precoding is used for data transmitted within two adjacent time slots by Zhang et al. [4][5]. This method can achieve full diversity gain and improve the system spectral efficiency, but increases the decoding complexity. The cyclic delay diversity method is introduced in the full-rate cooperative transmission model by Kwon et al. [6]. It reduces the system detection complexity and increases the channel frequency selectivity and the system frequency diversity as well. In system with CDD, the BER performance is affected by the cyclic delay value [7][8]. To realize the time diversity in the system with CDD further, the concept of randomized cyclic delay diversity was proposed by Plass et al. [9][10]. In the system with randomized cyclic delay diversity, the OFDM signals transmitted through each antenna are cyclically delayed in time domain respectively, where the cyclic delay value is selected randomly. This scheme not only obtains both the system frequency diversity and the system time diversity, but also improves the system BER performance without increasing the detection complexity at the receiving end. However, there exists a widespread problem that the system spectral efficiency is low. To improve the system diversity gain, the RCDD method is introduced to the multi-relay cooperative communication by Choi et al. [11][12], but the system cannot achieve full-rate data transmission and thus reduces the system spectrum efficiency.

In order to maximize the utilization of wireless network spectrum resources and improve the system diversity gain, this paper presents a distributed space-time-frequency coding scheme with randomized cyclic delay method and linear constellation precoding technology in the NAF full-rate multi-relay cooperative communication model.

The rest of the paper is organized as follows. Section II describes the NAF full-rate distributed cooperative communication model. The decoding scheme of the full-rate distributed space-time-frequency codes is given in Section III. Section IV analyses the performance of the proposed scheme. Finally, Section V presents the conclusion and future work.

II. SYSTEM MODEL

As shown in Fig. 1, the non-orthogonal amplify and forward (NAF) full-rate cooperative transmission model was used in this paper. It consists of a source node S, M relay nodes R_i, $i = 1, 2, ..., M$, and a destination node D. Each node is equipped with a single antenna, and operates in half-duplex mode. In the first time slot, the source node broadcasts the first signal to all the relay nodes and the destination node. In the second time slot, the source node continues to transmit the next signal to the destination node. Meanwhile the relay node forwards the signal received in the first time slot to the destination node after some processing. Source node completes the transmission of two symbols in two time slots, so that it can achieve full-rate transmission.

In order to overcome the "short-board effect", the symbol transmitted from the source node S is coded by linear constellation precoding. Assume that $\mathbf{S}_1$, $\mathbf{S}_2$ are the

 ISBN: 978-1-61208-305-6

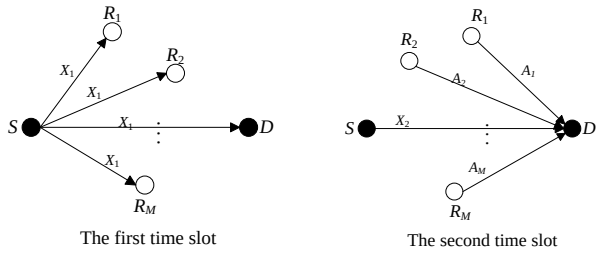


Figure 1. NAF full-rate cooperative transmission model.

symbols transmitted in two adjacent time slot from the source node in the frequency domain, respectively. $\mathbf{X}_1$, $\mathbf{X}_2$ are the symbols after linear constellation precoding. Where $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{X}_1$ and $\mathbf{X}_2$ are length-N_F column vectors, N_F represents the number of OFDM subcarriers. The relationship between $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{X}_1$, $\mathbf{X}_2$ is given as (1).

$$\begin{bmatrix} X_1(p) & X_2(p) \end{bmatrix}^T = \mathbf{\Theta} \begin{bmatrix} S_1(p) & S_2(p) \end{bmatrix}^T \tag{1}$$

where, $S_1(p)$ and $S_2(p)$ correspond to the data transmitted on p-th subcarrier of $\mathbf{S}_1$ and $\mathbf{S}_2$, $p \in [1, N_F - 1]$. $\mathbf{\Theta}$ is 2×2 linear constellation precoding matrix. In the first time slot, the source node S broadcasts $\mathbf{X}_1$ to all the relay nodes and the destination node D. Thus, the received signals at i-th relay node and destination node in p-th subcarrier are shown in (2) and (3).

$$R_i(p) = \sqrt{P_1} f_i(p) X_1(p) + N_i(p) \tag{2}$$

$$Z_1(p) = \sqrt{P_1} h(p) X_1(p) + W_1(p) \tag{3}$$

where, P_1 is the transmit power of the source node, $h(p)$ and $f_i(p)$ represent the complex channel fading coefficients of S to D and S to the i-th relay node R_i in the frequency domain, respectively. $X_1(p)$ is the data transmitted on the p-th subcarrier of X_1, $N_i(p)$ and $W_1(p)$ are both complex additive white Gaussian noise (AWGN) with zero-mean and variance N_0.

In the second time slot, the relay node R_i forwards the signal R_i to the destination node D after randomized cyclic delay. Thus, the signal transmitted from the relay node R_i is

$$A_i(p) = \alpha R_i(p)^{CDD} \tag{4}$$

To ensure the transmit power of the relay node, the amplifying power factor gain $\alpha = \sqrt{P_2 / (N_0 + P_1)}$, where P_2 is the transmit power of the relay node, $R_i(p)^{CDD}$ is $R_i(p)$ with random cyclic delay.

$$R_i(p)^{CDD} = R_i(p) e^{-j\frac{2\pi}{N_F} p \delta_i} \tag{5}$$

where, δ_i is the cyclic delay value. The randomized cyclic delay scheme is used in this paper, so δ_i is selected randomly between $[0, \ N_F - 1]$. Because delay in time domain is equivalent to the phase shift in the frequency domain, so in the frequency domain it is expressed as $e^{-j(2\pi p \Delta_i)/N_F}$.

The source node broadcast $\mathbf{X}_2$ to the destination node D in the second time slot as well. Thus, the received signal at the destination node in the p-th subcarrier is:

$$\begin{aligned} Z_2(p) = \alpha \sum_{i=1}^{M} \left(g_i(p) R_i(p)^{CDD} \right) + \\ \sqrt{P_3} h(p) X_2(p) + W_2(p) \end{aligned} \tag{6}$$

where, P_3 is the transmit power of the source node in the second time slot, $X_2(p)$ is the data transmitted on p-th subcarrier of $\mathbf{X}_2$, $g_i(p)$ is complex channel fading coefficient of the i-th relay node R_i to the destination node D in the frequency domain, $W_2(p)$ is complex additive white Gaussian noise with zero-mean and variance N_0.

By combining (5) and (6) we can obtain:

$$\begin{aligned} Z_2(p) = \alpha \sqrt{P_1} \sum_{i=1}^{M} \left(g_i(p) f_i(p) e^{-j\frac{2\pi}{N_F} p \delta_i} \right) X_1(p) \\ + \sqrt{P_3} h(p) X_2(p) + \tilde{W}_2(p) \end{aligned} \tag{7}$$

where,

$$\tilde{W}_2(p) = \alpha \sum_{i=1}^{M} g_i(p) N_i(p) e^{-j\frac{2\pi}{N_F} p \delta_i} + W_2(p) \tag{8}$$

For a large value of M, $\sum_{i=1}^{M} |g_i(p)|^2 \approx M$ with a high probability. So, there has the result

$$\mathbf{E}\left[\tilde{\mathbf{W}}_2 \tilde{\mathbf{W}}_2^H \right] \approx N_0 \left(1 + M\alpha^2 \right) \mathbf{I}_N \tag{9}$$

where $\tilde{\mathbf{W}}_2 = \left[\tilde{W}_2(0), \cdots, \tilde{W}_2(N_F - 1) \right]$, $(\bullet)^H$ denotes the complex transpose conjugate and $\mathbf{E}[\bullet]$ represents the expection operation. Let us set $K = N_0 \left(1 + M\alpha^2 \right)$. In order

to normalize the noises of the signals received in two time slots to be zero-mean and unit variance complex Gaussian, we divide (7) and (8) by $\sqrt{N_0}$ and $\sqrt{K}$ respectively. The normalized signals are given by (10) and (11).

$$\tilde{Z}_1(p) = \sqrt{c_1\rho}h(p)X_1(p) + \tilde{W}_1(p) \tag{10}$$

$$\begin{aligned}\tilde{Z}_2(p) = \sqrt{c_2\rho}\sum_{i=1}^{M}\left(g_i(p)f_i(p)e^{-j\frac{2\pi}{N_F}p\delta_i}\right)X_1(p) \\ + \sqrt{c_3\rho}h(p)X_2(p) + \tilde{V}(p)\end{aligned} \tag{11}$$

where $\tilde{W}_1(p)$ and $\tilde{V}(p)$ are both complex additive white Gaussian noise with zero-mean and variance N_0 . Let $\rho = P/N_0$ be the total transmit signal-to-noise ratio of the system, where $P = P_1 + MP_2 + P_3$. Finally, the power allocation coefficients c_1 , c_2 , c_3 are

$$c_1 = \frac{P_1}{P} \tag{12}$$

$$c_2 = \frac{P_1P_2}{P(\sum_{i=1}^{M}|g_i(p)|^2 \cdot P_2 + P_1 + N_0)} \tag{13}$$

$$c_3 = \frac{P_3(P_1 + N_0)}{P(\sum_{i=1}^{M}|g_i(p)|^2 \cdot P_2 + P_1 + N_0)} \tag{14}$$

III. DECODING

From (10) and (11), we can obtain that the received signals at the destination node on the *p*-th subcarrier are

$$\begin{aligned}Y(p) = \sqrt{\rho}\begin{bmatrix}\sqrt{c_1}X_1(p)h(p) \\ \sqrt{c_3}X_2(p)h(p) + \sqrt{c_2}X_1(p)\Psi(p)\end{bmatrix} \\ + \tilde{N}(p)\end{aligned} \tag{15}$$

where,

$$\Psi(p) = \sum_{i=1}^{M} g_i(p)f_i(p)e^{-j\frac{2\pi}{N_F}p\delta_i} \tag{16}$$

$$Y(p) = \begin{bmatrix}\tilde{Z}_1(p) & \tilde{Z}_2(p)\end{bmatrix}^T \tag{17}$$

$$\tilde{N}(p) = \begin{bmatrix}\tilde{W}_1(p) & \tilde{V}(p)\end{bmatrix}^T \tag{18}$$

Using (1), we can rewrite (15) as (19).

$$\begin{aligned}Y(p) &= \sqrt{\rho}\begin{bmatrix}\sqrt{c_1}h(p) & \\ \sqrt{c_2}\Psi(p) & \sqrt{c_3}h(p)\end{bmatrix}\begin{bmatrix}X_1(p) \\ X_2(p)\end{bmatrix} + \tilde{N}(p) \\ &= \sqrt{\rho}\begin{bmatrix}\sqrt{c_1}h(p) & \\ \sqrt{c_2}\Psi(p) & \sqrt{c_3}h(p)\end{bmatrix}\boldsymbol{\Theta}S(p) + \tilde{N}(p)\end{aligned} \tag{19}$$

where $S(p) = \begin{bmatrix}S_1(p) & S_2(p)\end{bmatrix}^T$, from the above equation, we can see that the received signals at the destination node are the linear transformation of the signals transmitted through the source node, that is to say, the proposed scheme has a one-dimensional equivalent channel, and with the number of the relay node increasing, the detecting complexity does not increase. Denote

$$\Gamma(p) = \sqrt{\rho}\begin{bmatrix}\sqrt{c_1}h(p) & \\ \sqrt{c_2}\Psi(p) & \sqrt{c_3}h(p)\end{bmatrix}\boldsymbol{\Theta} \tag{20}$$

Then, (19) is equivalent to

$$Y(p) = \Gamma(p)S(p) + \tilde{N}(p) \tag{21}$$

Assume that the Channel Status Information (CSI) is perfectly known at the destination node, the signals $\tilde{\mathbf{S}}$ transmitted through the source node can be detected by using the Maximum Likelihood (ML) detection method.

IV. SIMULATION RESULT AND ANALYSIS

The BER performance of the proposed scheme is simulated by Matlab in this section. Signals transmitted from the source node are firstly encoded by convolutional code and modulated into QPSK, where the rate of convolutional code is 1/2. Then, the signals are encoded with linear constellation precoding matrix $\boldsymbol{\Theta}$. The channel model is frequency selective fading channel with the carrier frequency of 3.5GHZ, and the multipath number is 2. The signal received at the relay node will be randomly cyclically delayed and then amplified and forward to destination node. The power allocation scheme in [4] was used in the paper. At the destination node, we use the ML detection method for decoding. The BER performance of system with RCDD and system with FCDD are shown in Fig. 2, Fig. 3 and Fig. 4, where the number of the relay node is 1, 2, 4, and the number of OFDM subcarriers is 32, 128 and 1024.

From Fig. 2, we can see that, in the case of small number of subcarriers, the BER performance of system with RCDD is substantially the same as the system with FCDD, and it improves as the number of the relay node increases.

From Fig. 3 and Fig. 4, we can see that, in the case of

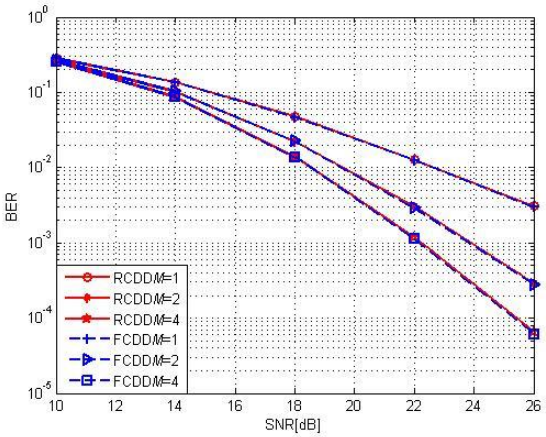


Figure 2. BER performance vs. SNR $N_F = 32$ and $M = 1, 2, 4$.

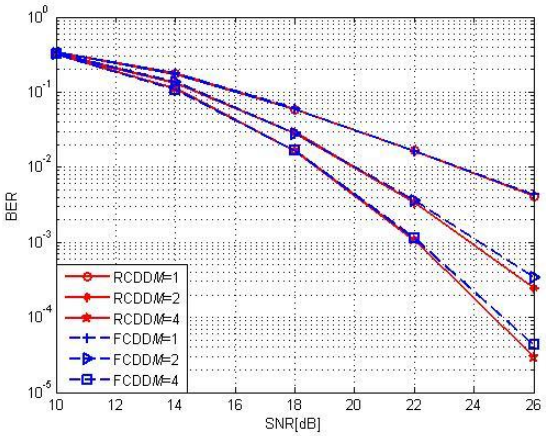


Figure 3. BER performance vs. SNR $N_F = 128$ and $M = 1, 2, 4$.

large number of subcarriers, the proposed scheme can achieve better BER performance. From Fig. 2, Fig. 3 and Fig. 4, we can see that randomized cyclic delay can excavate the system diversity gain further as the number of subcarriers increases.

V. CONCLUSION AND FUTURE WORK

Unlike fixed cyclic delay coding scheme used in NAF full-rate transmission system, the cyclic delay value is selected randomly in the proposed full-rate distributed space-time-frequency codes scheme, which is more practical than fixed cyclic delay. The proposed scheme is able to achieve better BER performance than the FCDD as the number of OFDM subcarriers increase, and it has the advantages of low detection complexity, i.e., the decoding complexity does not increase as the number of relay nodes increases.

In the future, we are planning to extend the present study to bi-directional distributed cooperative communication systems.

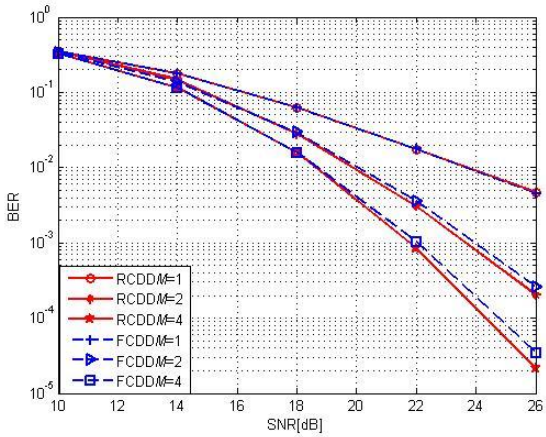


Figure 4. BER performance vs. SNR $N_F = 1024$ and $M = 1, 2, 4$.

ACKNOWLEDGMENT

Project supported by the National Natural Science Foundation of China (No. 60972040), the Anhui Provincial Natural Science Foundation (No. 11040606Q06), the Provincial Project of Natural Science Research for Colleges and Universities of Anhui Province of China (No. KJ2012A003), the PhD Start-up Foundation (No. 33190217) and the 211 Project of Anhui University.

REFERENCES

[1] X. Li, Q. Zhang, G. Zhang, and J. Qi, "Joint Power Allocation and Subcarrier Pairing for Cooperative OFDM AF Multi-Relay Networks," IEEE Communications Letters, vol.17, no.5, May 2013, pp. 872-875.

[2] L. J. Rodr ́guez, and N. H. Tran, A. Helmy, and T. Le-Ngoc, "Optimal Power Adaption for Cooperative AF Relaying with Channel Side Information," IEEE Transactions on Vehicular Technology, vol. pp, no.99, 2013, pp. 1-11.

[3] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading Relay Channels: Performance Limits and Space-Time Signal Design," IEEE Journal on Selected Areas in Communications, vol. 22, no. 6, Aug. 2004, pp. 1099-1109.

[4] W. Zhang, and K. B. Letaief, "Full-Rate Distributed Space-Time Codes for Cooperative Communication," IEEE Transactions on Wireless Communications, vol. 7, no. 7, Jul. 2008, pp. 2446-2451.

[5] H. Phan, T. Q. Duong, and H.-J. Zepernick, "Full-Rate Distributed Space-Time Coding for Bi-directional Cooperative Communication," IEEE 2010 5th International Symposium on Wireless Pervasive Computing (ISWPC), May 2010, pp.22-26.

[6] U. Kwon, C. Choi, and G. Im, " Full-Rate Cooperative Communications with Spatial Diversity for Half-Duplex Uplink Relay Channels, " IEEE Transactions on Wireless Communications, vol. 11, no. 8, Nov. 2009, pp. 5449-5454.

[7] Y.-J. Kim, H.-Y. Kim, M. Rim, and D.-Wo. Lim, "On the Optimal Cyclic Delay Value in Cyclic Delay Diversity," IEEE Trans. Broadcasting, vol. 55, no. 4, Dec. 2009, pp. 790-795.

[8] S.-H. Hur, M.-J. Rim, B. D. Rao, and J. R. Zeidler, "Determination of Cyclic Delay for CDD Utilizing RMS

Delay Spread in OFDMA Multiuser Scheduling Systems," IEEE 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Nov. 2010, pp. 506-510.

[9] S. Plass, A. Damman, G. Richter, and M. Bossert, "Resulting Channel Characteristics from Time-Varing Cyclic Delay Diversity in OFDM," 2007 IEEE 66th Vehicular Technology Conference, Sept. 2007, pp. 1336-1340.

[10] G. Richter, M. Bossert, E. Costa, and M. Weckerle, "On Time-Varing Cyclic Delay Diversity," European Transactions Telecommunications, vol. 17, no. 3, Mar. 2006, pp. 3100-3105.

[11] S. Choi, J.-H. Park, and D.-J. Park, "Randomized Cyclic Delay Code for Cooperative Communication System," IEEE, Communications Letters, vol. 12, no. 4, Apr. 2008, pp. 271-273.

[12] S. Choi, and D.-J. Park, "Performance of Randomized Cyclic Delay Code Encoded by Convolutional Coding," IEEE, Vehicular Technology Conference, May 2008, pp. 2370-2373.

 ISBN: 978-1-61208-305-6

EEMC-MAC: An Energy Efficient Protocol for Multi-Channel Wireless Networks

Thiago Fernandes Neves
Department of Computer Science
University of Brasilia, UnB
Brasilia, Brazil
e-mail: tfn.thiago@cic.unb.br

Jacir Luiz Bordim
Department of Computer Science
University of Brasilia, UnB
Brasilia, Brazil
e-mail: bordim@unb.br

***Abstract*—The popularization of wireless network technologies has driven the quest for efficient solutions in the use of the available resources. In particular, there is an increasing demand for solutions to reduce energy consumption and improve spectrum use. In this context, this work addresses the problems of energy efficient multi-channel assignment and communication scheduling in wireless networks. Considering that the channel allocation is an NP-complete problem, this paper presents a time and energy-efficient protocol. The protocol divides its operation in management and transmission stages. Empirical results show that the management stage, in average, takes less than 5% from the total protocol execution time, while the transmission stage is optimum in terms of energy consumption.**

***Keywords*—wireless networks; energy efficient protocols; multi-channel assignment; scheduling.**

I. INTRODUCTION

The quest for uninterrupted, high throughput wireless networks has been highly influenced by the popularisation of mobile devices and social networks. This trend in mobile applications has boosted the research efforts for Medium Access Control (MAC) protocols capable to cope with the demand. One of the major concerns in designing such protocols is to keep energy consumption at acceptable levels as the wireless devices are often powered by batteries. *Topology control* and *duty-cycle* are two energy saving strategies widely adopted in wireless networks [1]. Topology control techniques typically allow wireless devices to adjust their transmission power in order to conserve energy without affecting network connectivity [3]. Duty-cycle schemes allow wireless devices to alternate between inactive and active mode. When in active mode, the device is able to send or receive data and when in doze mode, the device is in energy conservation mode, where it is not able to send or receive data. This last strategy is particularly challenging as a device in doze mode is not able to receive data packets. Thus, the development of techniques to ensure that communicating devices will be active at the same time when there is data to send or receive are necessary [4].

The available MAC protocols are usually designed for single-channel environments [5]. Such protocols, especially in dense scenarios, have problems with packet collision, thus increasing packet retransmission, end-to-end delay and reducing throughput. Multiple communication channels have been used to increase throughput in wireless networks [6]. Such channels can be obtained via opportunistic spectrum access techniques, thus obtaining temporary access to unused licensed frequencies [7]. With the availability of multiple channels, Frequency Division Multiple Access (FDMA) based techniques, for example, allows to select several communication channels with non-overlapping and non-interfering frequencies. Thus, a pair of nodes can communicate at the same time and without interference since they are allocated to different channels.

A number of works consider the use of multiple channels in wireless networks [8], [9], [10]. Some of these works combined multi-channel MAC protocols with duty-cycle schemes to increase network throughput and decrease energy consumption. Tang *et al.* [11] proposed a multi-channel energy efficient protocol that minimizes energy consumption in wireless sensor networks. The proposed protocol allows the transmitting nodes to estimate the receiving node activation time without the use of a control channel. Incel *et al.* [8] proposed a multi-channel MAC protocol for wireless sensor networks. The proposed scheme works in a distributed fashion and makes communication schedule based on Time Division Multiple Access (TDMA) algorithms. This approach has been shown to reduce packet collision by informing the nodes what periods of time they need to be active. The proposed scheme, however, focus on maximising the throughout rather than minimizing energy consumption.

Zhang *et al.* [9] proposed a multi-channel MAC protocol for ad hoc networks. The proposed scheme works by dividing its operation in management and transmission stages. At the beginning of the management stage, all the nodes wishing to communicate turn to the control channel. The management stage dynamically adjusts its duration based on the traffic and it is used to allow the nodes to reserve data channels using the common control channel. During the transmission window, nodes communicate using several channels, while non-communicating nodes stay in doze mode. In previous work, we proposed an energy efficient protocol for multi-channel allocation and transmission scheduling in wireless networks, termed ECOA-BP [12]. As in [9], the ECOA-BP divides its operation in management and transmission stages and uses a control channel during the management stage. This technique uses efficient transmission assignment and duty-cycle strategy to alternate the nodes between active and inactive modes, thus reducing the power drainage rate.

The previous works show that is possible to reduce energy consumption at the cost of higher communication time. Conversely, one can minimise the communication time at the cost of higher energy consumption [13]. Clearly, there is a challenge in finding a compromise between these conflicting parameters. Both Zhang et al. [9] and Neves et al. [12] focus on balancing these parameters. However, the use of a single control channel in the management stage, independently of the number of available channels, can be a bottleneck, as it increases the communication time [2].

 ISBN: 978-1-61208-305-6

This paper addresses the problems of multi-channel allocation, transmission scheduling and energy consumption in wireless networks. As in related works, it is assumed that the devices work on batteries and have a single transceiver, capable of tuning to one of the several available channels and to switch between active (regular energy consumption) and inactive (reduced energy consumption) operation modes. The time is assumed to be slotted and its duration to be long enough to ensure a single data packet transmission or reception. In this context, this paper proposes a time and energy-efficient protocol capable of performing multi-channel allocation and transmission scheduling in a wireless setting. The proposed protocol operation is divided in management and transmission stages. Unlike most of the similar proposals, the proposed protocol uses all the available channels in both management and transmission stages. Empirical results show that the management stage, in average, takes less than 5% from the total protocol execution time, while the transmission stage is optimum in terms of energy consumption.

The remainder of this paper is organised as follows. Section II describes the considered communication model. Section III presents the channel assignment problem along with an energy-efficient heuristic to tackle it. Section IV presents the proposal, while Section V presents the empirical results. Section VI concludes the work.

II. Communication Model

An ad hoc network consists of a set of n nodes. A single-hop network setting can be represented by a complete graph G'_n, where each node in this network has a single transceiver and a unique identifier, that is known by the other nodes. The communication scenario of this network, on the other hand, can be represented by a directed graph $G = (V, E)$ (communication graph), where $V = \{v_1, v_2, ..., v_n\}$ is a set of nodes (vertices) and $E \subseteq V^2$ is a set of communications (edges). Consider $E = \{e_1, e_2, ..., e_p\}$, where $e_h = \{(v_s, v_d) | \{v_s, v_d\} \subseteq V, s \neq d\}$, $1 \leq h \leq p$, as a set of edges representing the communication graph of the network G'_n. Each edge $e_h = (v_s, v_d) \in E$ represents a communication between a source node v_s and a destination node v_d. There are no parallel edges between any two nodes. Consider s_i as the transmission set of a node v_i, which contains all the nodes that v_i ($v_i \in V$), has data packets to send, and d_i as the reception set of a node v_i, which contains all the nodes that have data packets to v_i. This way, each node v_i has $\tau_i = |s_i| + |d_i|$ data packets to send and receive.

As an example, Figure 1 represents a possible communication graph for a network topology G'_n. In this figure, $V = \{v_1, v_2, v_3, v_4\}$ and $E = \{e_1, e_2, e_3\}$, where $e_1 = (v_1, v_2)$, $e_2 = (v_1, v_4)$ and $e_3 = (v_3, v_2)$. In this communication graph, the node v_1 has data to send to nodes v_2 and v_4 , thus, $s_1 = \{v_2, v_4\}$, and no data to receive, thus $d_1 = \emptyset$. Similarly, $s_2 = \emptyset$, $d_2 = \{v_1, v_3\}$, $s_3 = \{v_2\}$, $d_3 = \emptyset$, $s_4 = \emptyset$ and $d_4 = \{v_1\}$.

As presented in [9], this paper assumes that data transmission/reception occur in time slots, with each transmission/reception taking exactly one time slot. In each time slot t_j, $j \geq 0$, where t_j is equal to the time interval $[t_j, t_{j+1})$, a node can be in active or inactive operation mode. When active, a node can send or receive data. Otherwise, the node can save power in the idle mode. That is, energy consumption is associated with the amount of time that the node remains in active mode. Consider $C = \{c_1, c_2, ..., c_k\}$ as the set of available channels for communication. When a channel $c_i, 1 \leq i \leq k$, is used by a pair of nodes in the time slot t_j, it will be unavailable for other nodes in this time slot. In the case that two or more transmitting nodes use the channel c_i during time slot t_j, a collision occurs and the data packets are lost.

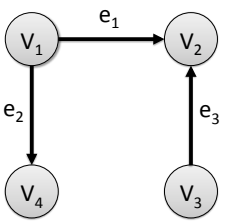


Figure 1: Communication graph example with 4 nodes.

III. The Channel Assignment Problem (CAP)

In a network environment, where many frequency channels are available, the task of channel assignment that satisfies the interference constraints and maximizes the throughput is known as the Channel Assignment Problem (CAP). To prevent interference between communications, the same channel cannot be allocated for two pairs of neighbouring nodes simultaneously. In its general form, the CAP problem is equivalent to the Generalised Graph-coloring Problem (GCP), which is known as a NP-complete problem [14]. Given the communication graph G and k channels in the presented communication model, the CAP consists in performing the communications using the minimum amount of time and communication channels. Note that if $k = 1$ this problem is simplified, once all the communications must be serialised. However, in the general case scenario, optimum solutions are complex to obtain.

Because of the NP-completeness of the CAP, many researchers proposed heuristics and approximation algorithms for the problem, which, however, can not guarantee optimum solutions. Proposed alternatives vary from neural networks, to genetic and graph theory based heuristics [14]. Next, an heuristic based on graph theory to solve the CAP problem is presented.

A. ECOH: An Edge Coloring Heuristic

Figure 2 presents an Edge Coloring Heuristic, termed ECOH. The proposed heuristic takes as input a communication graph $G = (V, E)$ and a number k of available channels. As output, the algorithm returns a list of communication sets, called CS. The list of communication sets is defined by $CS = \{CS_1, CS_2, ..., CS_r\}$, with $CS_i \subseteq E$ and the elements in CS_i are disjoint, $1 \leq i \leq r \leq |E|$. The basic idea behind the proposed heuristic is the distribution of edges belonging to E into r communication sets, so that the edges contained in a set CS_i have no dependencies with each other. The selection criterion is the choice of an edge belonging to

Algorithm ECOH(G,k)

1: $G = (V, E)$, $r \leftarrow 0$;
2: **while** ($E \neq \emptyset$) **do**
3: $r \leftarrow r + 1$;
4: Select an edge e of the vertex with higher degree in E;
5: $CS_r \leftarrow e$, $E \leftarrow E - e$;
6: **for** (each $e_h \in E$) **do**
7: **if** (no vertex in $e_h \in CS_r$) **and** ($|CS_r| \leq k$) **then**
8: $CS_r \leftarrow CS_r \bigcup e_h$;
9: $E \leftarrow E - e_h$;
10: **end if**
11: **end for**
12: **end while**
13: $CS \leftarrow \{CS_1, CS_2, ..., CS_r\}$;

Figure 2: The proposed edge colouring heuristic (ECOH).

a greater degree vertex in E. This edge will be part of the initial transmission set CS_i and it will be a comparison base for the other edges belonging to E. Only the edges without dependences with other elements in CS_i will be removed from E and incorporated into this set. An edge is considered not dependent on a set of edges when it does not share any vertex with the edges on this set. The procedure is repeated until the r transmission sets are formed and the set E is empty.

To better understand the operations of the ECOH, consider as input the communication graph represented in Figure 1 and the number of available channels to be equal to 2 ($k = 2$). Thus, $E = \{e_1, e_2, e_3\}$, where $e_1 = (v_1, v_2)$, $e_2 = (v_1, v_4)$ and $e_3 = (v_3, v_2)$. Suppose that the edge e_2 is inserted into the first set of edges in CS_1, line 5 (Figure 2). Going through all edges of E, line 6, the algorithm checks that the edge e_3 has no dependence on the set CS_1 and decides to insert it, line 8. As there are no more edges in E without dependencies with the elements of the set CS_1, the algorithm terminates the loop. A new loop is then started, line 2, and the variable r is incremented to 2. In the new loop, the algorithm inserts the edge e_1 in the set CS_2, ending the algorithm, since the condition $E = \emptyset$ is reached, line 2. In this example, o algorithm output would be $CS = \{CS_1, CS_2\}$, where $CS_1 = \{e_2, e_3\}$ and $CS_2 = \{e_1\}$. Note that, according to the algorithm, $|CS_i| \leq k$. That means each communication set has at most $k = 2$ disjoint elements. This construction allows the nodes in CS_i to communicate concurrently using the k channels in the same time slot.

B. ECOH: Involved Complexities

The ECOH heuristic has two main loops aligned: one that runs up to $E = \emptyset$ and another that compares vertices of edges in E with vertices in CS_i, looking for edges without dependencies. Thus, ECOH runs in $O(p^2)$ time, where $|E| = p$. Note that this complexity considers the worst case scenario where the nodes are not able to communicate in parallel or $k = 1$. For latter reference, consider the following result:

Lemma 1: Given a number of channels k and a communication graph $G = (V, E)$, the ECOH heuristic computes a list of communication sets $CS = \{CS_1, CS_2, ..., CS_r\}$ such that the edges in each set CS_i, $1 \leq i \leq r$ have no dependency with one another and $k \leq |CS_i|$. The ECOH computes the r lists in $O(p^2)$ operations.

IV. Proposed Protocol

This section presents the details of the proposed protocol, named Energy Efficient Multi-Channel MAC Protocol (EEMC-MAC Protocol). This protocol aims to perform multi-channel allocation and scheduling to enable data communication. In addition, the protocol performs these tasks in order to minimise both energy consumption and the time required to transmit data. First, it is presented some routines that are used in the protocol. Then, the protocol details are presented, followed by the protocol complexities.

A. Transmission Set Grouping Routines

Recall that each node $v_i \in V$ contains a set s_i, which identifies the destination nodes to which node v_i has data to send. In this subsection, the objective is combining such sets in a given node. The $CombineGroup$ routine, presented in Figure 3, aims to achieve this goal using a single communication channel. The routine takes as input: a set of nodes g_i, $g_i \subseteq V$, and a communication channel c_i. In the first step of the algorithm, each node in g_i computes a consecutive local ID, line 2. Let v_l be the node with the highest ID in g_i. The loop in lines $3-8$ combines the transmission sets s_j, $1 \leq j \leq l$ such that the local node v_l knows $s_l \cup s_{l-1} \cup ... \cup s_1$ in the end of the algorithm. Note that the above routine is very efficient in terms of energy consumption, once each node stays is active mode for just 2 time slots: one to send its transmission set and other to receive the transmission set from another node. Now, suppose that $|C| = k$, $k > 1$, channels are available, where C is the set of channels $C = \{c_1, c_2, ..., c_k\}$. In this case, the $CombineGroup$ routine could be improved to take advantage of several channels.

The routine $CombineTS$, depicted in Figure 4, shows how transmission sets can be combined, using multiple channels simultaneously. Similarly to the $CombineGroup$ routine, $CombineTS$ takes two input parameters: a group of nodes g_l, $g_l \subseteq V$, and a set of channels C, where $|g_l| = l$ and $|C| = k$. The routine is only executed if $k \geq \lfloor \frac{l}{2} \rfloor$, this way, all the transmissions in g_l can be parallelized in the k channels. At the beginning, all the active nodes compute their local ID in the range $[1, ..., l]$, line 4. The procedure grows a binary tree, combining the leaf nodes and working its way to the root using the k available channels, lines 5-13. In the end of the algorithm, the local node v_1 will have all the transmission sets $s_l \cup s_{l-1} \cup ... \cup s_1$. For latter reference, consider the following result:

Lemma 2: The $CombineGroup$ routine combines the transmissions sets in g_i in $|g_i| - 1$ time slots using a single channel with each node in active mode for 2 time slots. The $CombineTS$ routine combines the transmission sets in g_l in $\log k + 1$ time slots, using $|C| = k$ channels and with each node in active mode for at most $\log k + 1$ time slots, where $k \geq \lfloor \frac{l}{2} \rfloor$ and $|g_l| = l$. For both routines, it is assumed that each node can send at most 1 data packet to any other node in the network and it has a local buffer of l^2 bits.

Algorithm CombineGroup(g_i,c_i)

1: Let $|g_i| = l$;
2: Each node computes its local ID within the range $[1, ..., l]$ such that $g_i = \{v_1, v_2, ..., v_l\}$;
3: **for** $j \leftarrow 1$ **to** $l - 1$ **do**
4: Nodes v_j and v_{j+1} enter in active mode;
5: v_j sends its transmission set s_j to v_{j+1} using channel c_i;
6: Node v_{j+1} attaches s_j to s_{j+1};
7: Node v_j enters in inactive mode;
8: **end for**

Figure 3: Algorithm that combines the transmission sets in a group.

Algorithm CombineTS(g_i,C)

1: Let $|g_i| = l$ e $|C| = k$;
2: **if** ($k \geq \lfloor \frac{l}{2} \rfloor$) **then**
3: Let $C = \{c_1, c_2, ..., c_k\}$;
4: Each node computes its local ID within the range $[1, ..., l]$ such that $g_i = \{v_1, v_2, ..., v_l\}$;
5: **while** ($l > 1$) **do**
6: **for** ($i \leftarrow 0$ **to** ($\frac{l}{2} - 1$)) in parallel **do**
7: Assign channel c_{i+1} to pair (v_{i+1},v_{l-i});
8: v_{l-i} sends its transmission set s_{l-i} to v_{i+1};
9: v_{i+1} makes $s_{i+1} = s_{i+1} \bigcup s_{l-i}$;
10: v_{l-i} goes into inactive mode;
11: **end for**
12: $l \leftarrow l/2$;
13: **end while**
14: **end if**

Figure 4: Algorithm that combines the transmission on all groups.

B. EEMC-MAC Details

Next, the details of the EEMC-MAC protocol is presented. The EEMC-MAC is divided in two stages: management and transmission, which are described in the next subsections.

1) EEMC-MAC: Management Stage: The management stage main idea is to ensure that a leader node gets all the s_i transmission sets from all the nodes $v_i \in V$. This process must occur in a energy efficient way and use the maximum number of available channels. Then, the leader node can join all the communication sets and create the communication graph $G = (V, E)$. Figure 5 shows the management stage steps. In the beginning of the algorithm all the nodes are in inactive mode. If $k < \frac{n}{2}$, the n nodes in the set $V = \{v_1, v_2, ..., v_n\}$ are divided in k groups of nodes g_1, g_2, ..., g_k, lines 2-3. Once each node knows the values of k, n and its local ID, it has the condition to identify the group it belongs to. The goal is to reduce the number of active stations down to k. In the next step, k calls of the routine $CombineGroup$ are performed, line 5. As described above, the routine $CombineGroup$ will combine the transmission sets in each group g_i to just one node per group and the other nodes involved are set to inactive mode. The routine $CombineTS$ is called for all the active nodes. This routine will guarantee that all the transmission sets will be combined and forwarded to a single node $v_m \in V$, lines 9-10. Node v_m will hold all the network transmission sets. At the end, node v_m uses the transmission sets information to build the communication graph $G = (V, E)$, line 11.

Algorithm ManagementStage(n, k)

1: All the nodes in $V = \{v_1, v_2, ..., v_n\}$ start in inactive mode;
2: **if** ($k < \lfloor \frac{n}{2} \rfloor$) **then**
3: Divide the nodes in V into k groups: $g_1, g_2, ...g_k$;
4: **for** $i \leftarrow 1$ **to** k in parallel **do**
5: Execute *CombineGroup*(g_i,c_i);
6: **end for**
7: **end if**
8: Let g_l denote de set of active stations;
9: The active stations execute *CombineTS*(g_l,C);
10: Let v_m be the last active station from the previous step;
11: Node v_m uses the transmission sets information to build the communication graph G;

Figure 5: Building the communication graph from the obtained transmission sets.

Algorithm TransmissionStage

1: Let v_m be the network node leader (from the previous stage) with the communication graph G;
2: Node v_m executes $ECOH(G, k)$ and gets the communication sets $CS = \{CS_1, CS_2, ..., CS_r\}$;
3: All the nodes in V enter in active mode and tunes into channel c_1. Node v_m broadcasts CS in channel c_1. All the nodes in V receives the CS broadcast and enters in inactive mode;
4: **for** $i \leftarrow 1$ **to** r **do**
5: **for** $j \leftarrow 1$ **to** $|CS_i|$ in parallel **do**
6: Select an unused edge $e_h = \{v_s, v_d\}$ from CS_i;
7: Nodes v_s and v_d enter in active mode;
8: Node v_s sends a packet to v_d using channel c_j;
9: Nodes v_s and v_d enter in inactive mode;
10: Mark the edge e_h from CS_i as used;
11: **end for**
12: **end for**

Figure 6: Each node proceeds to the assigned channel to transmit and receive data packets.

Depending on the input, the $ManagementStage$ may call the $CombineGroup$ routine in parallel for k groups, taking $\frac{n}{k} - 1$ time slots for execution with each node staying in active mode for 2 time slots. The $CombineTS$ routine is called once and takes $\log k + 1$ time slots to be executed with each node v_i staying in active mode for τ_i time slots. Considering the results in Lemma 2, the following result is presented:

Lemma 3: Given a set of nodes V and a set of channels C, where $|C| = k$ and $|V| = n$, the $ManagementStage$ combines the transmission sets in $O(\frac{n}{k} + \log k)$ time slots with each node in active mode for $O(\log k)$ time slots.

2) EEMC-MAC: Transmission Stage: The transmission stage of the EEMC-MAC protocol begins immediately after the management stage. In this stage, the leader node v_m already computed the communication graph G. Figure 6 presents the $TransmissionStage$ details. In the beginning of the algorithm, the leader node v_m has the communication graph of the entire network G. To solve the communication dependences, the leader node v_m executes the ECOH heuristic and gets the list of communication sets $CS = \{CS_1, CS_2, ..., CS_r\}$, lines 1-2. The ECOH ensures that $|CS_i| \leq k$, that is, each set has

at most the number of available channels and all the elements in each set CS_i are disjoint. In a next step, all the nodes enter in active mode and tune into channel c_1 to receive the CS broadcast from the leader node v_m and then return into inactive mode, line 3. The first loop, line 4-12, goes from 1 to r (the number of communication sets) and the second loop goes from 1 to the number of elements in the communication set indicated in the previous loop, lines 5-11. It begins by selecting an unused edge from the set CS_i. The nodes in this set enter in active mode, line 7, and tune in the indicated channel and perform the data transmission, line 8, returning to inactive mode after the transmission, line 9. This process continues until all the nodes in each communication set exchange their data sets.

The EEMC-MAC transmission stage runtime depends on how the ECOH heuristic creates the list of communication sets $CS = \{CS_1, CS_2, ..., CS_r\}$. Analysing the $TransmissionStage$ (Figure 6), note that r time slots are necessary to perform all the transmissions represented in CS. The transmissions in CS_i are disjoint and can be performed concurrently using multiple channels. An additional time slot is used for the CS broadcast. Thus, in the worst case scenario, where every transmission has to be serialised, r would be equal to the number of edges in the communication graph, that is, $r = p$. However, in the best case scenario, all the transmission in G could be spread over the k available channels, that is, $r = \frac{p}{k}$. This way, the transmission stage total runtime is between $\Omega(\frac{p}{k})$ and $O(p)$ time slots per execution. It is considered that the ECOH heuristic execution and CS diffusion can be performed in the same time slot. It should be noted that every node in this stage, except when the nodes received the CS broadcast, enters in active mode only to send or receive data. This way, the energy consumption for each node $v_i \in V$ in this stage is equal to $\tau_i + 1$. The EEMC-MAC transmission stage complexities are summarised below:

Lemma 4: The $TransmissionStage$ takes $\Omega(\frac{p}{k})$ and $O(p)$ time slots per execution with each network node $v_i \in V$ in active mode for no more than $\tau_i + 1$ time slots.

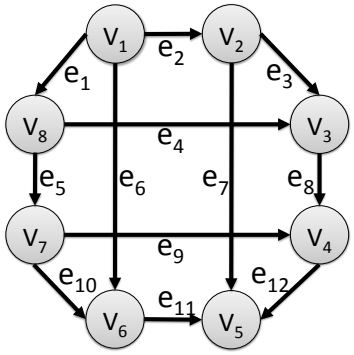
C. EEMC-MAC: Main Procedure and Complexities

The main procedure of the EEMC-MAC protocol executes the two aforementioned stages in sequence. Thus, the EEMC-MAC total runtime is $\Omega(\log k + \lceil \frac{p}{k} \rceil)$ and $O(\lceil \frac{n}{k} \rceil + \log k + p)$ time slots. Theorem 1 summarises the protocol complexities, considering the Lemmas 3 and 4.

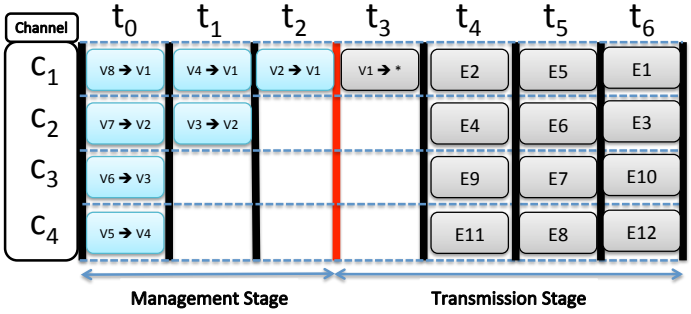
Theorem 1: The EEMC-MAC protocol solves the multi-channel medium access control and transmission scheduling in a time slotted, synchronized, single hop wireless settings, represented by a communication graph $G = (V, E)$, in $O(\lceil \frac{n}{k} \rceil + \log k + p)$ time slots, with each node $v_i \in V$ in active mode for $O(\log k + \tau_i)$ time slots, where $|V| = n$, $|E| = p$, $|C| = k$ and τ_i is the number of data packets that node v_i has to send and receive.

D. EEMC-MAC: A working example

To exemplify the protocol application, consider the communication graph represented by Figure 7a. This graph has 8 vertices, $V = \{v_1, v_2, ..., v_8\}$, and 12 edges, $E = \{e_1, e_2, ..., e_{12}\}$. Consider the presence of $k = 4$ communication channels.



(a)



(b)

Figure 7: (a) Communication graph example with 8 nodes. (b) Channel representation for the EEMC-MAC protocol.

Figure 7b represents a possible data transmissions using 4 channels, the proposed communication graph and the EEMC-MAC protocol. The protocol main procedure begins with the execution of the management stage (shown in Figure 5). Once the number of channels is large enough ($k \geq \lfloor \frac{n}{2} \rfloor$), the routine $CombineTS$ is called. This routine will group all the transmission sets s_i of nodes in V, using the $k = 4$ channels, until the leader node v_1 gets all the communication sets, represented in time slots t_0 to t_2 in Figure 7b. This procedure of grouping transmission sets ends the management stage. The transmission stage (shown in Figure 6) starts immediately after the management stage ends. In this stage, the leader node v_1 uses the ECOH heuristic (Figure 2) to solve the graph communication dependencies and to obtain the list of communication sets CS. This list allows to perform the transmission scheduling, containing the channel and time slot each node must tune to send or receive data. Note that the ECOH heuristic ensures that parallel transmission does not share vertices in common. The leader node, then, broadcasts CS to all the other nodes in time slot t_3. Time slots t_4 to t_6 represent the scheduled packet transmissions.

V. Simulation

The evaluation of the proposed protocol has been performed through simulation. For this purpose, a simulator has been developed in Matlab environment [15]. The simulator incorporates the characteristics of the EEMC-MAC protocol, described in Section IV. To verify the goodness of the proposed solution, the simulation results are compared with the optimum

 ISBN: 978-1-61208-305-6

solutions. This section describes the simulation parameters, the evaluation metrics and then presents the obtained results.

A. Simulation Parameters and Evaluation Metrics

The simulation has been conduced for a varying number of nodes, data packets per node and data channels. The number n of nodes assume the following values: $8, 16, 32, 64, 128, 256$. The number of data packets per node assume values in one of the five different ranges: 0% to 20%, 21% to 40%, 41% to 60%, 61% to 80% and 81% to 100%. Each range represents a percentage of the maximum number of transmissions per node. Recall, from the communication model, that each node can have a maximum of $n-1$ outcoming edges, that is, a node can send 0 or 1 packets to any destination in the communication graph per EEMC-MAC execution cycle. For example, in a setting with 16 nodes using the first range (0% to 20%), each node would have from 0 to $20\% * (16-1) = 3$ data packets to send. The number of channels is defined as 2^i, with i going from 0 to $\lfloor \frac{n}{2} \rfloor$. The simulation results are drawn from the average of 200 simulation runs for each setting.

The protocol execution time will be evaluated considering: (i) the percentage of time the protocol spend in the transmission stage (R_{ts}); and (ii) the ratio between the protocol transmission stage time and the optimum transmission stage time (R_{opt}). The R_{ts} is defined as follows:

$$R_{ts} = \frac{T_t}{T_t + T_m}, \quad (1)$$

where T_m is the number of time slots the protocol needed in the management stage, T_t is the number of time slots needed in the transmission stage. Note that lower R_{ts} value indicates that the protocol incurs in a lower message overhead to transmit the data items. The R_{opt} is defined similarly:

$$R_{opt} = \frac{T_t}{T'_t}, \quad (2)$$

where T'_t is the optimum transmission stage time (in time slots). The R_{opt} values indicates the gap between the current transmission stage time and the optimum time. Clearly, when $R_{opt} = 1$, the EEMC-MAC protocol achieved the minimum necessary time to complete the transmission stage.

B. Simulation Results

Figures 8a and 8c present the simulation results for R_{ts}, considering $n = 16, 32, 64$ nodes and 5 different transmission configurations, that is, a communication graph with 0% to 20% of maximum number of edges, 21% to 40%, and so on. The x-axis shows the number of channels while in the y-axis presents the R_{ts} values.

It can be observed that the R_{ts} values decrease with an increase in the number of channels. This was expected as an increase in the number of channels allows for a larger number of parallel transmissions, decreasing the time needed for the transmission stage. As the number of nodes increase, the management stage time decreases. This can be observed in Figure 8c, where the R_{ts} is close to 100%. That is, the protocol spends most of its time in the transmission stage.

As can be observed, the percentage of time the protocol needs for management is minimal when compared with the total protocol execution time. In fact, this time was, in average, less than 5% from the total protocol execution time. In what follows, a closer look is taken at the time needed for the transmission stage.

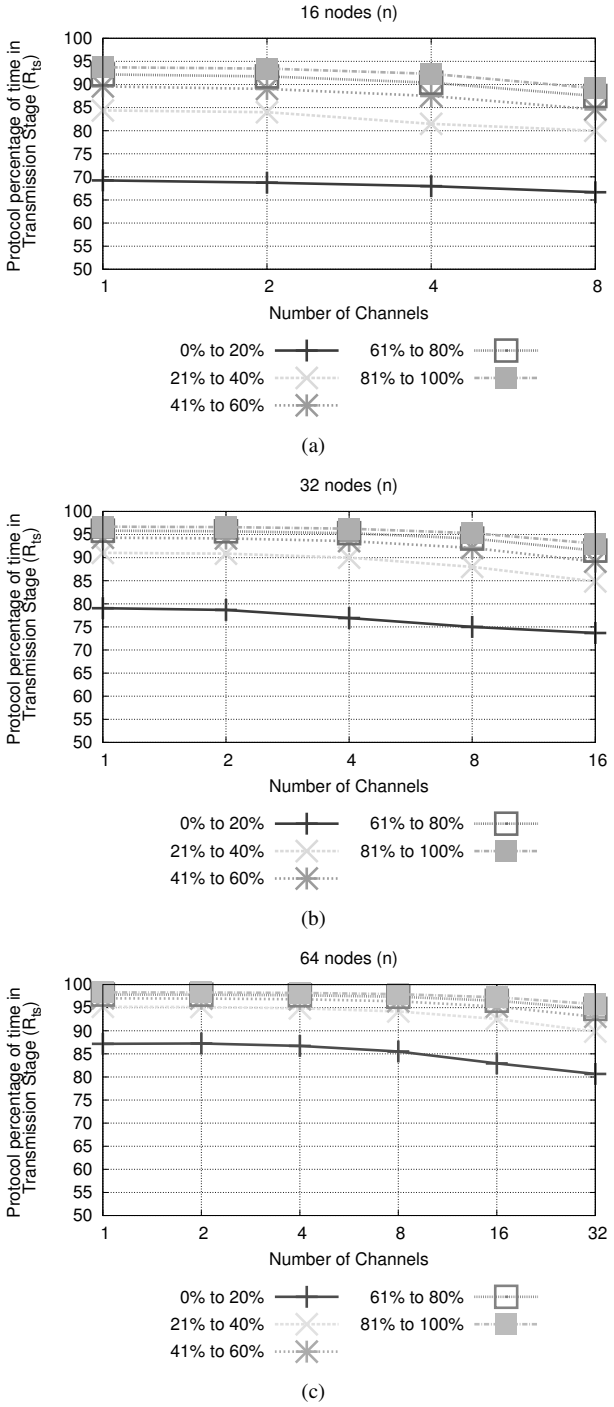


Figure 8: Simulation results for $n = 16, 32, 64$ nodes and metric R_{ts}.

Figure 9 presents the simulation results for the metric R_{opt}. From the Vizing theorem [16], it is a valid lower bound to assume that the optimum channel assignment execution time is equal to $\Delta(G)$, where $\Delta(G)$ is the maximum graph degree. Thus, for comparison purpose, it is assumed that $T_{t'} = \Delta(G)$.

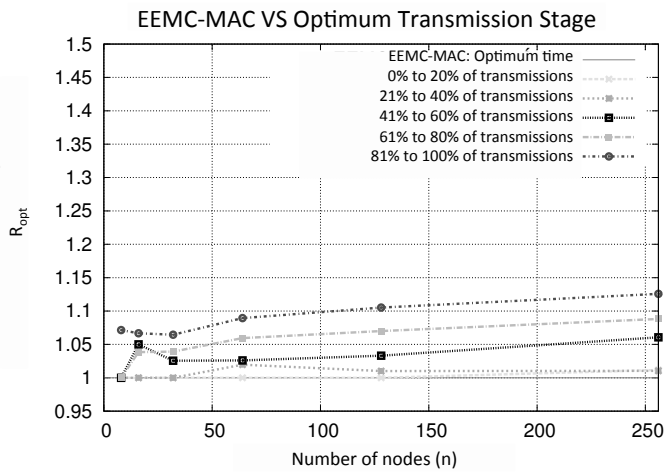


Figure 9: Simulation results for metric R_{opt}.

In the x-axis in Figure 9 shows the number of nodes in the communication graph while in the y-axis presents the R_{opt} values. The number of data items per node follows the ranges defined previously.

It can be observed in Figure 9 that $R_{opt} \approx 1$ when a lower packet load (first two ranges) is presented. The R_{opt} values increase with the number the of nodes and transmissions per communication graph. However, even in such cases, the EEMC-MAC transmission stage execution time was always less than 15% higher when compared with the optimum transmission stage time. A larger communication graph increases the number of similar choices in the selection criterion of the protocol transmission scheduling. Which, in turn, increases the chance of producing unfavourable scheduling, increasing the communication time. Note that the choice of an inappropriate transmission scheduling at a given step S impacts in the choice of other transmissions at step $S + 1$. From the results for metric R_{opt} it is concluded that the EEMC-MAC achieved performance close to the optimum in many cases. When the average of all the communication setting is computed, the EEMC-MAC is less than 5% from the optimum time.

VI. Conclusion

The increasing popularization of mobile devices and the emergence of high content applications, increased the need for high throughput and energy efficient protocols for wireless networks. In this context, this work proposes an energy efficient protocol, named EEMC-MAC, for multi-channel allocation and transmission scheduling in wireless networks. The EEMC-MAC protocol divides its operation in management and transmission stages. The energy expenditure in the management stage is minimum and empirical results shows that this stage represents less than 5% of the total protocol operation time. The transmission stage is optimum in energy consumption and, when compared with the optimum transmission stage time, the protocol needs, in average, 4% more time. In future works, it is intended to address fault tolerance and to improve the communication model.

References

[1] P. Mohapatra and S. V. Krishnamurthy, “Ad Hoc Networks - Technologies and Protocols". Springer Sciente + Business Media, Inc. chapters 1, 3, 6, 2005.

[2] M. F. Caetano ; Lourenço, B. F. ; Bordim, J.L.,“On Performance Of the IEEE 802.11 in a Multi-channel Environment". In: International Conference on Computer Communications and Networks, 2013, Nassau.

[3] Y. Zhu, M. Huang, S. Chen, and Y. Wang, “Energy-efficient topology control in cooperative ad hoc networks,” Parallel and Distributed Systems, IEEE Transactions on, vol. 23, no. 8, 2012, pp. 1480–1491.

[4] K. Chowdhury, N. Nandiraju, D. Cavalcanti, and D. Agrawal, “CMAC - A multi-channel energy efficient MAC for wireless sensor networks,” in Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE, vol. 2., 2006, pp. 1172–1177.

[5] S. Tsao and C. Huang, “A survey of energy efficient mac protocols for IEEE 802.11 wlan,” Computer Communications, vol. 34, no. 1, 2011, pp. 54–67.

[6] Q. Zhao and B. M. Sadler, “A survey of dynamic spectrum access,” Signal Processing Magazine, IEEE, vol. 24, no. 3, 2007, pp. 79–89.

[7] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, “IEEE 802.22: The first cognitive radio wireless regional area network standard,” IEEE Stands in Communications and Networking, January 2009, pp. 130 – 138.

[8] O. Incel, L. Van Hoesel, P. Jansen, and P. Havinga, “MC-LMAC: A multi-channel mac protocol for wireless sensor networks,” Ad Hoc Networks, vol. 9, no. 1, 2011, pp. 73–94.

[9] J. Zhang, G. Zhou, C. Huang, S. Son, and J. Stankovic, “Tmmac: An energy efficient multi-channel mac protocol for ad hoc networks,” in Communications, 2007. ICC’07. IEEE International Conference on. IEEE, 2007, pp. 3554–3561.

[10] A. Raniwala, K. Gopalan, and T. Chiueh, “Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks,” ACM SIGMOBILE Mobile Computing and Communications Review, vol. 8, no. 2, 2004, pp. 50–65.

[11] L. Tang, Y. Sun, O. Gurewitz, and D. Johnson, “Em-mac: a dynamic multichannel energy-efficient mac protocol for wireless sensor networks,” in Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing. ACM, 2011, p. 23.

[12] T. F. Neves, M. F. Caetano, and J. L. Bordim, “An energy-optimum and communication-time efficient protocol for allocation, scheduling and routing in wireless networks,” in Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International. IEEE, 2012, pp. 848–854.

[13] J. Bordim, J. Cui, and K. Nakano, “Randomized time-and energy-optimal routing in single-hop, single-channel radio networks,” IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, vol. 86, no. 5, 2003, pp. 1103–1112.

[14] G. Audhya, K. Sinha, S. Ghosh, and B. Sinha, “A survey on the channel assignment problem in wireless networks,” Wireless Communications and Mobile Computing, vol. 11, no. 5, 2011, pp. 583–609.

[15] M. Grant, S. Boyd, and Y. Ye, “Cvx: Matlab software for disciplined convex programming,” 2008.

[16] V. G. Vizing, “On an estimate of the chromatic class of a p-graph,” Diskret. Analiz, vol. 3, no. 7, 1964, pp. 25–30.

[17] O. Younis and S. Fahmy, “HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks,” Mobile Computing, IEEE Transactions on ,vol. 3, no. 4, 2004, pp. 366–379.

 ISBN: 978-1-61208-305-6

Distance Estimation of Smart Device using Bluetooth

Joonyoung Jung, Dongoh Kang, Changseok Bae
Personal Computing Platform Research Team
Electronics and Telecommunications Research Institute
Deajeon, Korea
{jyjung21, dongoh, csbae}@etri.re.kr

Abstract— **Distance estimation identifies the distance between two machines in wireless network. The Received Signal Strength Indication (RSSI) of Bluetooth can be used to estimate distance between smart devices. The characteristic of Bluetooth RSSI value is different as environments. So, we have tested the relation between distance and Bluetooth RSSI value in several environments, such as indoor hall, meeting room, and ElectroMagnetic Compatibility (EMC) chamber environment. This paper shows the distance characteristic of Bluetooth RSSI from these experiment results. There are a lot of measurement errors at Bluetooth RSSI raw data. The minimum RSSI value is -88 dBm and the maximum RSSI value is -66 dBm at 11m of the indoor hall environment. The difference between maximum value and minimum value is 22 dBm. So, it is hard to estimate the distance using Bluetooth RSSI raw data. Therefore, we use the Low Pass Filter (LPF) for reducing the measurement errors. The minimum RSSI value is -80.6 dBm and the maximum RSSI value is -71 dBm in the same environment. The difference between maximum value and minimum value is just 8.4 dBm. The measurement error is significantly reduced. We compare the distance estimation between the Bluetooth RSSI raw data and LPF data at the EMC environment. This paper shows that the distance estimation is possible with small error rates using Bluetooth RSSI LPF data.**

Keywords-Distance Estimation; Bluetooth; RSSI

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are one of the essential research domains. There are many applications for WSNs in military and civil applications [1]. The Machine to Machine (M2M) distance estimation is a fundamental issue for a lot of applications of indoor WSNs, such as a Bluetooth and Zigbee. Distance estimation identifies the distance between two machines in wireless network. Such estimates are an important component of systems' localization, because they are used by the position computation and localization algorithm components. Different methods, such as RSSI, Time of Arrival (ToA), and Time Difference of Arrival (TDoA), can be used to estimate a M2M distance. Nowadays, lots of location systems have tried to estimate M2M distance using different models in wireless networks. For example, the Active Badge System used an infrared signal [2]. Cricket, developed at MIT, uses TDoA method [3]. Global Positioning System (GPS) uses ToA [4]. RADAR, developed at Microsoft, uses RSSI to estimate M2M distance [5]. SpotON is a RSSI-based ad-hoc localization system [6]. In this paper, we discuss the M2M distance estimation using Bluetooth RSSI.

The rest of the paper is organized as follows. Section II describes related work. In Section III, we describe distance characteristic of Bluetooth RSSI. In Section IV, we describe Bluetooth RSSI using a low pass filter. Section V provides the experimental results, and some concluding remarks are finally given in Section VI.

II. RELATED WORK

A. RSSI

RSSI can be used to estimate the M2M distance based on the received signal strength from another machine. The longer the distance to the receiver machine, the lesser the signal strength at received machine. Theoretically, the signal strength is inversely proportional to squared distance, and there is a known radio propagation model that is used to convert the signal strength into distance. However, in real environments, it is hard to measure distance using RSSI because of noises, obstacles, and the type of antenna. In these cases, it is common to make a system calibration [7], where values of RSSI and distances are evaluated ahead of time in a controlled environment. The advantage of this method is its low cost, because most receivers can estimate the received signal strength. The disadvantage is that it is affected by noise and interference. So, distance estimation may have inaccuracies. Some experiments [8] show errors from 2 to 3 m in some scenarios. Distance estimation using RSSI in real-world applications is still questionable because of inaccuracy [9]. However, RSSI could become the most used technology of distance estimation from the cost/precision viewpoint because of low cost [10]. A. Awad et al. [1] discuss and analyze intensively some approaches relying on the received signal strength indicator. The most important factor for proper distance estimation is to choose a transmission power according to the relevant distances. It was showed that even for noisy indoor environments an average positioning error of 50cm on an area of 3.5 x 4.5 m is possible by choosing the RF and algorithm parameters carefully based on empirical studies. S. Feldmann et al. [11] also presented an indoor positioning system based on signal strength measurements, which were approximated by the received RSSI in a mobile device. The functional dependence between the received RSSI and the distance was achieved by a well fitted polynomial approximation.

 ISBN: 978-1-61208-305-6

B. ToA

In ToA, the M2M distance is directly proportional to the time the signal takes to propagate from one machine to another, as shown in Fig. 1 [12]. ToA needs precisely synchronized machines.

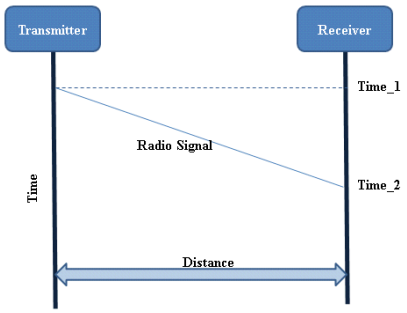


Figure 1. ToA distance estimation method

The distance between two machines is proportional to the signal transmitted time. That is, if a transmitter sends a signal at time time_1 and a receiver receives the signal at time time_2, the distance between transmitter and receiver is $d = P_r(time\_2 - time\_1)$, where P_r is the propagation speed of the radio signal, and time_1 and time_2 are the times when the signal was transmitted and received. S. Schwarzer et al. [13] presented a concept to measure the distance between two IEEE 802.15.4 compliant devices using ToA. It shows that compared to signal correlation in time, the phase processing technique yields an accuracy improvement of roughly an order of magnitude.

C. TDoA

TDoA is based on the difference in the times at which multiple signals from a single machine arrive at another machine. The machines must have extra hardware for sending two types of signals simultaneously, as shown in Fig. 2. These signals must have different propagation speeds, like RF and ultrasound. N. Priyantha et al. [14] presented a TDoA method using different propagation speed signals, like radio/ultrasound. K. Whitehouse et al. [7] used radio/acoustic signals. Usually, the first signal propagation speed is light, while the second signal has slower propagation speed. The second signal is six orders of magnitude slower than the first signal.

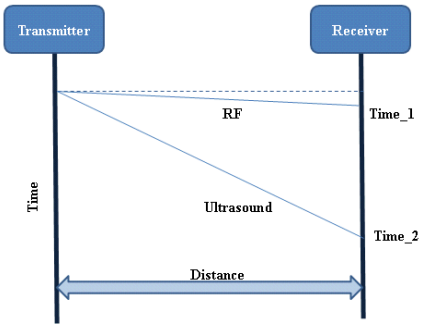


Figure 2. TDoA distance estimation method

An example of TDoA suitable for WSNs is used in [8] and depicted in Fig. 2, where the ultrasound pulse and radio signal are sent simultaneously. A receiver machine computes the difference time of the two signals. The distance can now be computed by the formula $d = (P_r - P_u)*(time\_2 - time\_1)$, where P_r and P_u are the propagation speed of the radio signal and ultrasound pulse, and time_1 and time_2 are the received times of the radio signal and ultrasound pulse, respectively. Another different and interesting way of computing distance among machines using the TDoA is proposed by Fu et al. [15], and is based on the Direct Sequence Spread Spectrum (DSSS) modulation technique. The distance estimation errors using TDoA are several centimeters. Experiments error with ultrasound performed in [8] is about 3 cm, where M2M distance was 3 m. In [16], the experiments error with acoustic sound is about 23 cm, where M2M distance was 2 m. TDoA system has precise distance estimation accuracy. However, it also has disadvantages. It needs extra hardware to send the second signal, which increases cost. And it has limited range of the second signal, which is about between 3 and 10 m according as transmitter power. To save a cost, Y. Fukuju et al. [17] presented a TDoA system that reduces configuration cost.

D. Location System using M2M Distance Estimation

The Active Badge System finds location information using an infrared signal [2]. Each person wears a small infrared badge. The badge sends a unique packet periodically or on demand. A server receives badge data using fixed infrared sensors in building and gathers this data. The Active Badge system provides absolute location information using this infrared distance information. Infrared signal has an effective range of several meters because of diffusion. Therefore, infrared signal range is limited to small or medium rooms. As mentioned above, drawbacks are limited range of infrared sensors and usage of diffused infrared in fluorescent lighting or direct sunlight.

Cricket uses TDoA method [3]. M2M distance error is about 3 cm, but this causes a huge burden on the receiver machine due to distributed computation and processing of ultrasound pulses and RF signal. The Cricket Location Support System finds location information using ultrasound pulse and RF signal. The RF signal is used for synchronization of the time measurement. Cricket estimates distance using TDoA and then finds location information using distance information. However, Cricket does not require a grid of ceiling sensors with fixed locations because its mobile receivers perform the timing and computation functions. A receiver receives multiple beacons, so it triangulates its position. Cricket has advantages that it has privacy and decentralized scalability. It also has disadvantages that it does not have centralized management and more it has the computational and power burden for timing and processing both the ultrasound pulses and RF signal on the mobile receivers.

RADAR was developed at Microsoft and used RSSI to estimate M2M distance [5]. It is based on an 802.11 Wireless LAN. A building-wide tracking system based on the IEEE 802.11 LAN wireless networking technology. RADAR

 ISBN: 978-1-61208-305-6

measures the signal strength and signal-to-noise ratio at the base station, and then it computes the position within a building using these data. RADAR's scene-analysis implementation has position error within about 3 meters with 50 percent probability, while the signal strength lateration implementation has position error about 4.3-meter with 50 percent probability.

SpotON is a RSSI-based ad-hoc localization system [6]. The SpotON system implements ad-hoc lateration with low cost tags. SpotON tags estimate distance between tags using radio signal attenuation. It can be used for relative and absolute position determination. In an ad-hoc location system, all of the machines become mobile machines with the same sensors and capabilities. To estimate their locations, machines cooperate with other nearby machines by sharing RSSI data. Machines in the cluster are located relative to one another or absolutely if some machines in the cluster have known locations. The techniques for building ad-hoc systems include triangulation, scene analysis, or proximity. Location sensing with ad-hoc infrastructure has a high scalability.

III. DISTANCE CHARACTERISTIC OF BLUETOOTH RSSI

We tested the relation between distance and Bluetooth RSSI in indoor hall, meeting room, and EMC chamber environment. These experiment results show the distance characteristic of Bluetooth RSSI in several environments

A. *Indoor hall*

We have measured Bluetooth RSSI with a notebook and a Nexus 7 in indoor hall environment as shown in Fig. 3. We measured 200 samples at each meter from 0m to 15m.

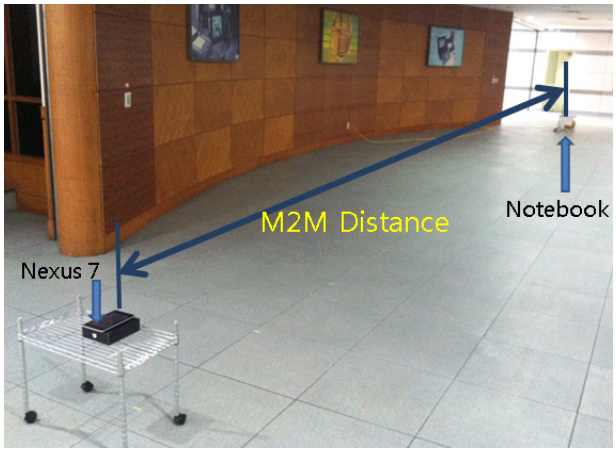
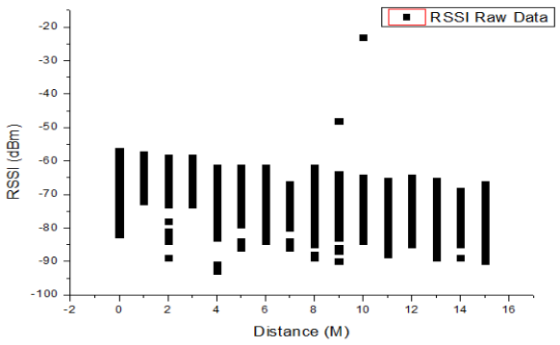
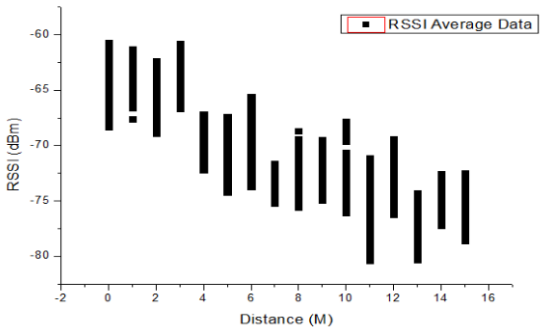


Figure 3. Bluetooth RSSI mesurement test in indoor hall environment

The result of these experiments is shown in Fig. 4. The RSSI raw data and average data are shown in Fig. 4 (a) and Fig. 4 (b). The RSSI average data average 10 RSSI raw data. The distance estimation is impossible with the RSSI raw data. However, the distance estimation may be possible coarsely with the RSSI average data. The RSSI average value is similar from 0m to 3m, from 4m to 6m, and from 7m to 15m.



(a) RSSI raw data



(b) RSSI average data

Figure 4. The Bluetooth RSSI measurement result in indoor hall environment

B. *Meeting Room*

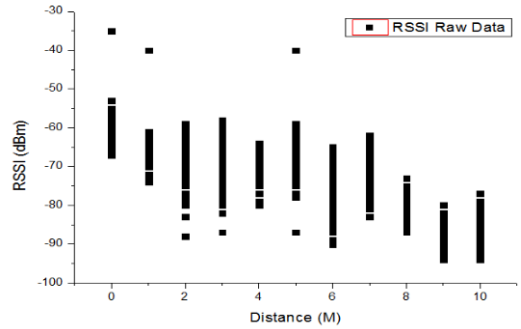
We have measured Bluetooth RSSI with a notebook and a Nexus 7 in meeting room environment, as shown in Fig. 5. We measured 200 samples at each meter from 0m to 10m. The meeting room door is located between 7m and 8m from the notebook. When Bluetooth RSSI value is measured from 8m to 10m, the meeting room door is closed.

Figure 5. Bluetooth RSSI mesurement test in meeting room environment

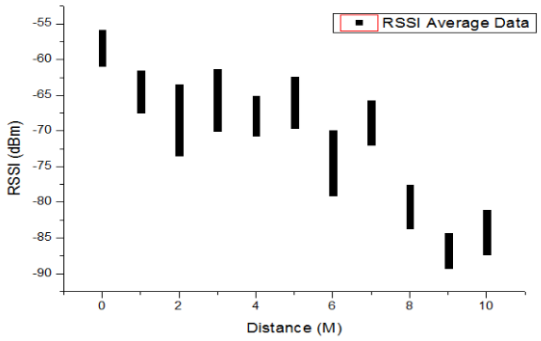
The RSSI raw data and the RSSI average data are shown in Fig. 6 (a) and Fig. 6 (b), respectively. It is hard to classify into inside and outside of the meeting room with the RSSI raw data. However, the minimum RSSI average value from 0m to 7m is -78.6 dBm and the maximum RSSI average value from 8m to 10m is -77.9 dBm. So, it may be possible

 ISBN: 978-1-61208-305-6

to classify into inside and outside of the meeting room with the RSSI average data.



(a) RSSI raw data



(b) RSSI average data

Figure 6. The Bluetooth RSSI mesurement result in meeting room environment

C. *EMC Chamber*

We have measured Bluetooth RSSI with a notebook and a Nexus 7 in EMC chamber environment as shown in Fig. 7. We measured 200 samples at each meter from 0m to 15m.

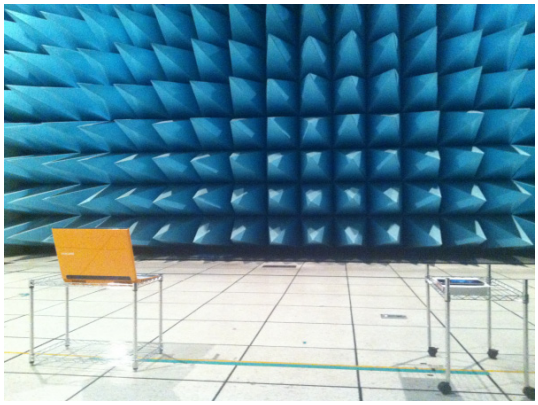
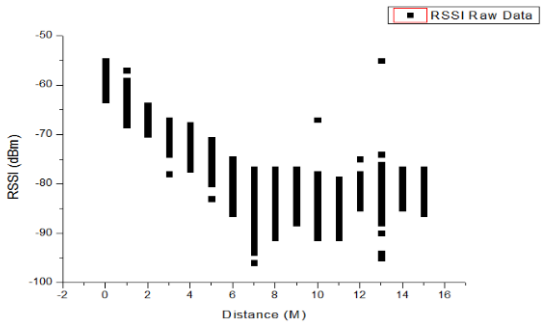
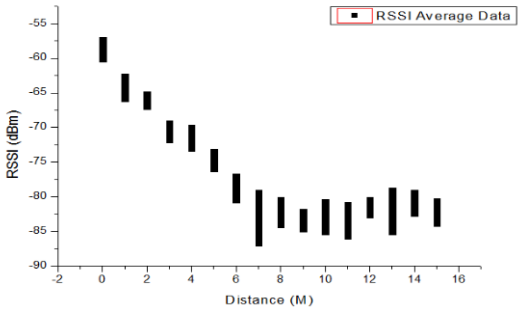


Figure 7. Bluetooth RSSI mesurement test in EMC chamber environment

The RSSI raw data and the RSSI average data are shown in Fig. 8 (a) and Fig. 8 (b), respectively.



(a) RSSI raw data



(b) RSSI average data

Figure 8. The Bluetooth RSSI mesurement result in EMC chamber environment

The RSSI value decreases linearly from 0m to 7m and has similar value from 7m to 15m. The distance estimation is performed well from 0m to 7m using the RSSI average value in EMC chamber environment.

IV. BLUETOOTH RSSI USING LOW PASS FILTER

The Bluetooth RSSI raw data at 11m of indoor hall environment is shown in Fig. 9. There are a lot of measurement errors at Bluetooth RSSI raw data. The minimum RSSI value is -88 dBm and the maximum RSSI value is -66 dBm. The difference between maximum value and minimum value is 22 dBm. So, it is hard to estimate the distance using Bluetooth RSSI raw data.

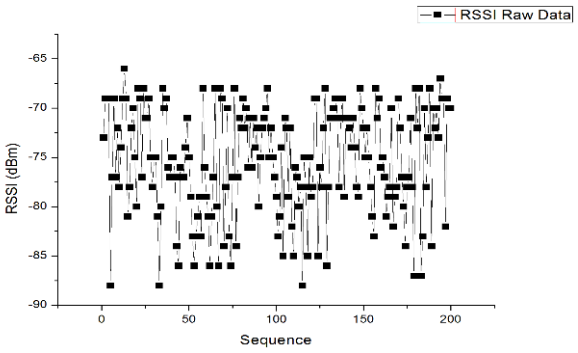


Figure 9. The Bluetooth RSSI raw data at 11m of indoor hall environment

 ISBN: 978-1-61208-305-6

LPF equation is (1). When the LPF is used, the RSSI deviation is reduced as seen Fig. 10, where (α = 0.8).

$$P_n = \alpha P_{n-1} + (1-\alpha) T_n \ . \quad (1)$$

The received signal strength (T_n) is the RSSI value receiving from the other smart device at k. And the LPF value (P_n) is the RSSI value of LPF at k. The constant (α) has a value that is bigger than 0 and lower than 1.

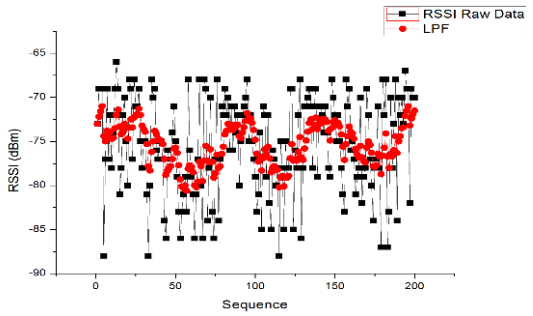


Figure 10. The Bluetooth RSSI LPF data at 11m of indoor hall environment

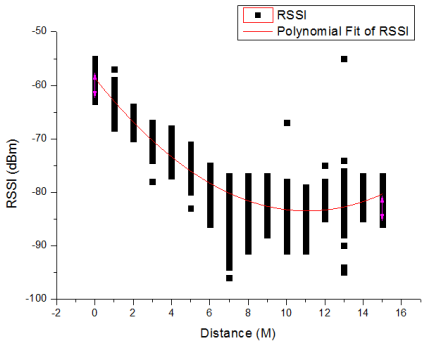
The minimum RSSI value is -80.6 dBm and the maximum RSSI value is -71 dBm. The measurement errors are significantly reduced. The difference between maximum value and minimum value is just 8.4 dBm.

V. EXPERIMENT RESULT

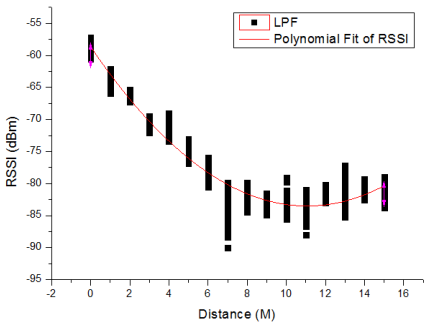
A. *Distance estimation comparision in EMC chamber*

We compare the distance estimation between the Bluetooth RSSI raw data and LPF data. The Bluetooth RSSI values are measured from 0m to 15m in EMC chamber.

Polynomial Regression (order 2) is used to estimate the distance as shown in Fig. 11. The R-square value of RSSI raw data is 0.867 and the standard deviation is shown in Fig. 12. The maximum value and the minimum value of standard deviation are 4.94 dBm and 1.26 dBm. The R-square value of RSSI LPF data is 0.958, which is better than R-square value of RSSI raw data significantly. The standard deviation is shown in Fig. 12. It shows also better result than those of RSSI raw data. The maximum value is 1.77 dBm and the minimum value is 0.36 dBm.



(a) Regression with RSSI raw data



(b) Regression with RSSI LPF data

Figure 11. Distance estimation comparision between RSSI raw data and LPF data in EMC chamber environment

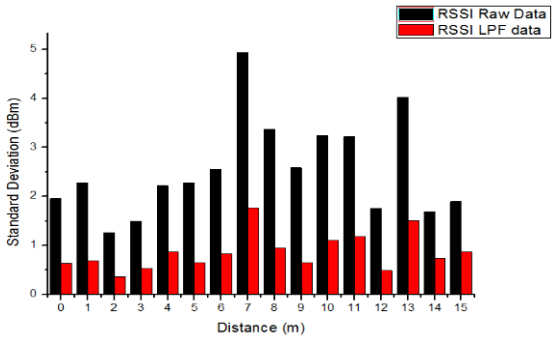


Figure 12. Standard deviation comparision between RSSI raw data and LPF data in EMC chamber environment

B. *Distance estimation application in meeting room*

One device is located in the meeting room statically and the other device is moved around inside and outside of the meeting room. The distance from 0m to 5m is inside of the meeting room and the distance from 6m to 10m is outside of the meeting room.

The standard deviation of RSSI value is too broad at each meter to estimate distance when the RSSI raw data are used, as shown in Fig. 13. So, we cannot distinguish between inside and outside of the meeting room.

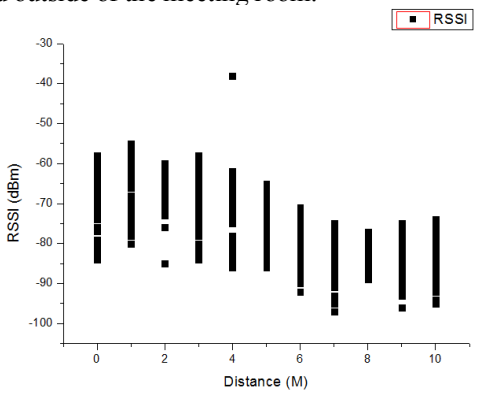


Figure 13. The RSSI raw data for distinguish between inside and outside of the meeting room

 ISBN: 978-1-61208-305-6

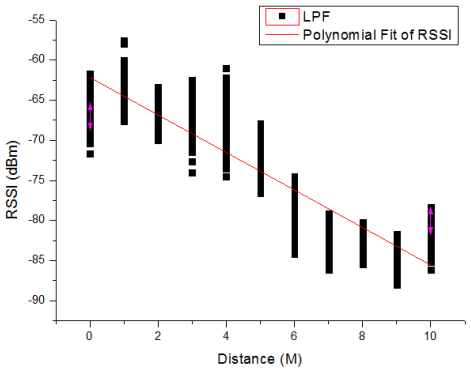


Figure 14. The RSSI LPF data for distinguish between inside and outside of the meeting room

The RSSI LPF data of the RSSI raw data are shown in Fig. 14 (α : 0.8). The R-square value is 0.838. The minimum RSSI value from 0m to 5m is -76.6dBm and the maximum RSSI value from 6m to 10m is -74.5dBm. The overlap region is only 2.1dBm. So, we can distinguish between inside and outside of the meeting room.

VI. Conclusion And Future Work

This paper addressed the distance estimation method and distance characteristic of Bluetooth RSSI. The distance estimation is impossible with the RSSI raw and may be possible coarsely with the RSSI average data in indoor hall environment. In meeting room environment, it is hard to classify into inside and outside of the meeting room with the RSSI raw data and may be possible to classify into inside and outside of the meeting room with the RSSI average data. The RSSI value decreases linearly from 0m to 7m and has similar value from 7m to 15m in EMC chamber environment. This paper considers the LPF for reducing the measurement errors. The measurement errors are significantly reduced. We compare the distance estimation between the Bluetooth RSSI raw data and LPF data at EMC chamber environment. With RSSI raw data, the R-square value is 0.867 and the maximum standard deviation value is 4.94 dBm. With RSSI LPF data, the R-square value is 0.958 and the maximum standard deviation value is 1.77 dBm. The LPF data shows better result than RSSI raw data. However, LPF data need to be improved for estimating distance more exactly. So, we will design a new algorithm to estimate distance with Bluetooth RSSI.

Acknowledgment

This work was supported by the IT R&D program of MKE/KEIT, [K10041801, Zero Configuration Type Device Interaction Technology using Device Sociality between Heterogeneous Devices].

References

[1] A. Awad, T. Frunzke, and F. Dressler, “Adaptive Distance Estimation and Localization in WSN using RSSI Measures,” 10th Euromicro Conference on Digital System Design Architectures, Methods and Tools, Aug. 2007, pp. 471-478.

[2] R. Want, A. Hopper, V. Falcao, and J. Gibbons., “The Active Badge Location System,” ACM Transactions on Information Systems, vol. 40, no. 1, Jan. 1992, pp. 91-102.

[3] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The Cricket Location-Support System,” Mobile Comp. and Networking, Aug. 2000, pp. 32–43.

[4] P. Enge and P. Misra, “Special Issue on GPS: The Global positioning System,” Proc. of the IEEE, vol. 87, no. 1, Jan. 1999, pp. 3-15.

[5] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," Proceedings of IEEE Infocom 2000, vol. 2, Mar. 2000, pp. 75-84.

[6] J. Hightower, C. Vakili, G. Borriello, and R. Want, "Design and Calibration of the SpotON Ad-Hoc Location Sensing System," University of Washington, Seattle, Technical Report UW CSE 01-08, Aug. 2001.

[7] K. Whitehouse and D. Culler, “Calibration as Parameter Estimation in Sensor Networks,” Wireless Sensor Networks and Apps., 2002, pp. 59-67.

[8] A. Savvides, C.-C. Han, and M. B. Strivastava, “Dynamic Fine-Grained Localization in Ad-Hoc Networks of Sensors,” 7th ACM/IEEE Int’l. Conf. Mobile Computing and Networking, Rome, Italy, 2001, pp. 166–179.

[9] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, “Range-Free Localization Schemes for Large Scale Sensor Networks,” MobiCom ’03, ACM Press, 2003, pp. 81-95.

[10] J. Bachrach and C. Taylor, “Localization in Sensor Networks,” Handbook of Sensor Networks: Algorithms and Architectures, I. Stojmenovic, Ed., Wiley, Sept. 2005.

[11] S. Feldmann, K. Kyamakya, A. Zapater, and Z. Lue, “An indoor Bluetooth-based positioning system: concept, implementation and experimental evaluation,” International Conference on Wireless Networks, 2003, pp. 109–113.

[12] B. Hofmann-Wellenho, H. Lichtenegger, and J. Collins, “Global Positioning System: Theory and Practice,” Springer-Verlag, 1997.

[13] S. Schwarzer, M. Vossiek, M. Pichler, and A. Stelzer, “Precise Distance Measurement with IEEE 802.15.4 (ZigBee) Devices,” IEEE Radio and Wireless Symposium, Mar. 2008, pp. 779-782.

[14] N. B. Priyantha, A. K. L. Miu, H. Balakrishnan, and S. Teller, “The Cricket Compass for Context-Aware Mobile Applications,” 7th ACM Int’l. Conf. Mobile Computing and Networking, Jul. 2001, pp. 1-14.

[15] Y. Fu et al., “The Localization of Wireless Sensor Network Nodes Based on DSSS,” Electro/Infor. Tech., 2006 IEEE Int’l. Conf., May 2006, pp. 465–469.

[16] K. Whitehouse, “The Design of Calamari: An Ad Hoc Localization System for Sensor Networks,” Master’s thesis, UC Berkeley, 2002.

[17] Y. Fukuju, M. Minami, H. Morikawa, and T. Aoyama, “DOLPHIN: An Autonomous Indoor Positioning System in Ubiquitous Computing Environment,” Workshop on Software Technologies for Future Embedded and Ubiquitous Systems, 2003, pp. 53-56.

 ISBN: 978-1-61208-305-6

Verification of Microwave Air-Bridging for Sky-Net

Chin E. Lin
Department of Aeronautics and Astronautics
National Cheng Kung University
Tainan 701, Taiwan, R.O.C.
e-mail: chinelin@mail.ncku.edu.tw

Ying-Chi Huang
Department of Aeronautics and Astronautics
National Cheng Kung University
Tainan 701, Taiwan, R.O.C.
e-mail: p48991191@mail.ncku.edu.tw

***Abstract*—The Sky Net project proposes a mobile base transceiver station (BTS) for mobile communication service from airborne. In the preliminary study, a Microwave Air-Bridging (MAB) system is established on an Ultra-Light Aircraft (ULA) to offer relay mobile communication. In order to maintain sufficient Quality of Service (QoS) data-link, microwave system needs to establish critical Line-of-Sight (LoS) antenna alignment between airborne to ground to assure polarization matching. The proposed GPS-to-GPS tracking algorithm uses GPS and barometric data by considering earth curvature to estimate azimuth and elevation angles for the ground dual-axis tracking mechanism. With high accuracy demand in tracking, all sensor data are pre-calibrated by determining their covariance. An additional Data Variance Filter (DVF) is applied to eliminate the sudden data error and white noise from sensors to avoid mechanism vibration and maintain accurate tracking response to match with ULA flight path. The goal tries to reduce the Bit Error Rate (BER) of MAB to accomplish accurate antenna alignment for microwave transmission in Sky-Net application.**

Keywords-Antenna Alignment; Microwave Air Bridging; Dynamic Filter; Dual Axis Motor Control.

I. INTRODUCTION

Microwave Air Bridging (MAB) has been widely used in telecommunication relaying recently. Without hardware limitations, MAB can easily be implemented to establish communication links from remote areas. Antenna alignment in MAB is the fundamental issue to solve before minimum Quality-of-Service (QoS) can be made for communication link. Most remote Base Transceiver Stations (BTSs) apply MAB application to connect wide spread mobile communication service network.

In disaster situation, some mobile BTSs might be flooded or destroyed to disable from communication services. A draft idea of Sky Net was proposed to create a BTS carrying on a high altitude Unmanned Aerial Vehicle (UAV) for mobile communication services. After some preliminary surveys, the first proposal using repeater is abandoned. Since both donor and service antennas use the same frequency, it is not able to eliminate antenna isolation problem on a UAV. A second proposal arose to use microwave link from airborne to ground.

The High Altitude Platform (HAP) or Remote Airborne Platform (RAP) for communication relay have been introduced and widely studied in the past few years. The concept of relaying the communication signal, basically, has two kinds of system architectures, such as Direct Relay (P2P) and Multi-Node Relay. In the Line-of-Sight (LOS) field, Unmanned Aerial Vehicle (UAV) can carry airborne transceiver or repeater to establish the P2P communication link between UAV and ground base [1] [2]. However, in general condition, the area needs telecommunication service, usually, has obstacle between the region and ground base. Using multi-UAV flying around the specific area to establish the communication link network is one of the ways to overcome this problem. Also, even the signal transmission range and the UAV collision problem can be well avoided [3] [4], the multi-UAV relay system still has some problem to overcome such as signal transmission latency after the multi-node network and the uncertainty of wireless link quality due to each UAV's actual flight path and real-time altitude.

The airborne BTS concept was introduced and accomplished by Wypych et al. [5]; however, the service was limited by MS-to-MS communication in service area, which means the user cannot get contact with people far away still. In this paper, the goal is establishing a microwave link system between airborne BTS and ground BTS, which connect to the backbone network of the telecom company.

With proper antenna alignment tracking, the exchange of telecommunication signal between terminals should meet the minimal acceptable BER in mobile BTS. In this phase of study, the airborne terminal is mounted on a ULA carrying the microwave bridge working at 5.8 GHz frequency band in Multiple Input and Multiple Output (MIMO) system with adaptive modulation. The proposed MAB system configuration is shown in Figure 1, as the preliminary idea.

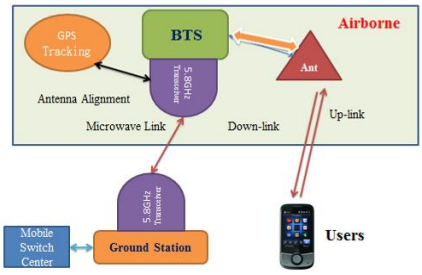


Figure 1. Microwave Air-Bridging system configuration for Sky-Net.

In general, in wireless applications, like UAV surveillance with real-time image or video [6], the transceiver, usually, works simply in one way transmission.

 ISBN: 978-1-61208-305-6

However, the microwave bridge needs both terminals to maintain a good signal strength and steady data-link quality in order to achieve high bandwidth data transmission. Once the Received Signal Strength Indicator (RSSI) drops below the threshold of transceiver specification, the link is disconnected. Antenna tracking control may be deterministic if the main lobe on both side terminal antennas is aligned or not. In general, -3dB is set from the pick of antenna pattern to define the Vertical Beamwidth Angle (VBA) and Horizontal Beamwidth Angle (HBA). This shows critical constraint of the antenna's directional characteristic. Usually, the threshold is carefully monitored such that the antenna can work normally.

Antenna misalignment will cause high bit error rate and make microwave bridge crash. The keys to maintain well antenna alignment are twofold: antenna polarization keeps perfect matching and both VBA and HBA of two side terminal antennas maintain overlapping. The theoretical foundation of antenna characteristic is used to describe further automatic tracking system design for antennas. To accomplish the antenna alignment, the microcontroller tracking mechanism and control algorithm are designed using the Cortex-M3 32-bit microcontroller and implemented on a dual axis mechanism for precision control.

In this paper, the concept of the antenna tracking system is introduced in Section II, part A. The GPS-to-GPS tracking method and the whole hardware architecture are well describe and also the data filter that is designed according to data variance is explained in part B. Section III shows how the stepper motor control strategy works to eliminate the platform vibration which could lead to antenna misalignment. Dynamic tests of tracking algorithm for antenna alignment are presented for microwave air bridging verification by open field flight experiments in Section IV.

II. TRACKING ALGORITHM

A. *GPS-to-GPS Tracking Method*

The existing microwave automatic antenna tracking device generally uses signal strength gradient to aiming at the target. The device will rotate azimuth angle to get the Received Signal Level (RSL) as high as possible and make signal gradient close to zero when it is aiming at the target. The elevation angle tracking works correspondingly. When the airborne terminal is relatively far away from the ground terminal, the signal will be diluted because of space propagation and signal multi-path effect to RSL simultaneously. The tracking performance worsens to result in lost alignment between the MAB terminals. Based on the Global Positioning System (GPS), a simple but effective method termed as GPS-to-GPS (G2G) tracking algorithm is proposed. The proposed G2G method receives GPS position information from two sites, i.e., airborne and ground, and transforms them into a specific coordinate system for alignment. For GPS information, both airborne and ground terminals will receive their latitude, longitude and altitude in Mean Sea Level (MSL), continually. The algorithm in the control core will compare their differences on each side and calculate the appropriate geometric azimuth and elevation angels between the airborne terminal and the ground terminal.

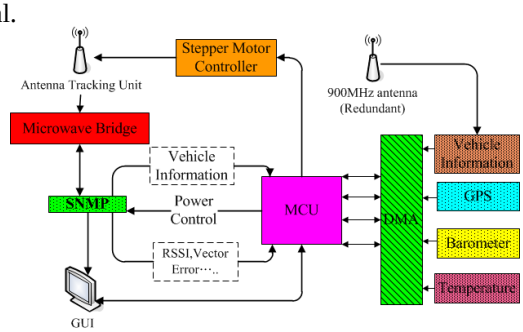


Figure 2. Backbone antenna tracking control system.

The backbone of antenna tracking control is shown in Figure 2. A dual-core mother board is fabricated to control the 2-axis rotation platform. The vehicle information of the airborne terminal is transmitted via microwave data link and Simple Network Management Protocol (SNMP) into MCU. A 900MHz wireless module is used as redundant link to ensure the integrity and reliability of the tracking control. All the data and command are transmitted by Direct Memory Access (DMA) to reduce CPU burden.

Coordinate transform plays a key part in the tracking control. The first step calculates and transforms the Longitude-Latitude-Altitude (LLA) information into an appropriate coordinate system to offer user with distance of East-to-West and North-to-South direction. The coordinate transform of LLA in WGS84 [7] into TM2 in TWD97 [8] can be rewritten by latitude and longitude in degree into East-North direction distance in meters. For the azimuth angle, as long as North and East direction can be obtained, the arctangent calculation can be applied to get a correct value. But for elevation angle, the situation is quite different. On the microwave transmission, the most concern on the VBA of both sides lies on their well overlap or not. Therefore, the elevation angle accuracy of mechanism unit becomes much important and sensitive than the azimuth angle. Figure 3 shows the difference between the calculations of two methods. The first method simply assumes the Cartesian coordinate system to ground terminal as the origin. The second method considers the Earth curvature error into calculation.

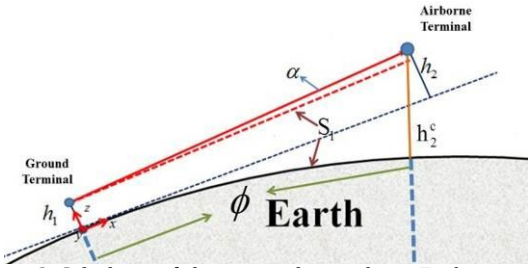


Figure 3. Calculation of elevation angle considering Earth curvature.

In Figure 3, if the airborne terminal and ground terminal are close enough, then the calculation can be simplified into a 3-axis Cartesian coordinate plane. As the coordinate being transformed into TM2 in TWD97 system, the distance S_1 between ground and airborne terminal can be obtained.

 ISBN: 978-1-61208-305-6

Meanwhile, using MSL for altitudes on both terminals as h_1 and h_2, the elevation angle α from the ground terminal can be expressed as:

$$\alpha = \arctan(\frac{h_2 - h_1}{S_1}) \quad (1)$$

However, as S_1 is far from both terminals, the curvature of earth will evidently affect the accuracy of elevation angle α. Figure 3 also shows how the curvature makes the error bigger as S_1 increasing. In TWD97 datum, the Earth model is an ellipsoid and the Earth radius scale is relatively huge comparing to ULA flight path. Within a small region, the Earth is reasonably assumed as a sphere not an ellipsoid. Therefore in figure 3, between the ground and the airborne terminals, the Earth radius is R and the incline angle between them toward Earth center is ϕ. The elevation angle α of the ground terminal can be recalculated as:

$$d_1 = R^2 + (h_2^c + R)^2 - 2R(h_2^c + R)\phi \quad (2)$$

$$d_2 = (h_1 + R)^2 + (h_2^c + R)^2 - 2(h_1 + R)(h_2^c + R)\cos\phi \quad (3)$$

$$\alpha = \cos^{-1}(\frac{h_1^2 + d_1^2 - d_2^2}{2 - h_1 d_1}) \quad (4)$$

B. Data Variance Filter

All the data from sensors should be filtered to eliminate noise or interference to cause unpredictable data error. Generally, software for Low-Pass Filter (LPF) or Complementary Filter [9, 10, 11, 12] can easily be adopted to compensate the unreasonable error or eliminate the sensor white noise. However, the time constant value or weight of the applied filter considerably affects the system response. In this paper, the Data Variance Filter (DVF) is adopted to add the variance of each sensor into the filter to generate the variable weight for LPF and avoid the internal data disturbance effect on the control output. The covariance values of GPS and barometric altitudes are also included to determine the time constant in complementary filter to get better altitude resolution.

DVF uses a simple equation to get the processed data at time y_t with data from t-1 and the data x acquired at time t. The weight is calculated by function ***Varf (α, β)*** and variance gain *kv*.

$$y_t = Varf(\alpha, \beta) y_{t-1} + kvx \quad (5)$$

The pre-calibrated step of DVF is getting the individual variance σ of sensors. Assuming that data have a Gaussian distribution, the sensor is implemented steady. The three variance derivations, α, β, and γ are generated as below:

$$\beta = 2\int_0^{\alpha} \frac{1}{\sqrt{2\pi}} \exp(-\frac{\tau^2}{2}) d\tau \quad (6)$$

$$\gamma = \left| \frac{x - y_{t-1}}{y_{t-1}} \right| \quad (7)$$

$$\alpha = \left| \frac{x - y_{t-1}}{\sigma} \right| \quad (8)$$

Equations (6) and (7) are determining data probability of each time data update. For Gaussian distribution, 99.73% of data will fall into the region of 3σ. Since the data sampling rate is 50 Hz in the proposed system, it is assumed that barometric and GPS data between two sampling times will not changing drastically. In order to reduce the calculation load on microcontroller, the 3σ data difference is set as the threshold to generate appropriate ***Varf(α, β)*** value from the following equation.

$$Varf(\alpha, \beta) = \beta - u(\alpha - 3)\{\beta - kv\} \quad (9)$$

The variance deviation will be calculated in each computation loop, and the variance gain (*kv*) will be automatically changed in each iteration.

$$kv = \gamma - u(\alpha - 1)\{\gamma - 1\} \quad (10)$$

The important part of using DVF falls into how to get the correct variance for each sensor. As the sampling frequency is relatively higher than system characteristic, the static data can be used to get the variance to fit for the real condition. Figure 4 shows how the variance affects the result of DVF.

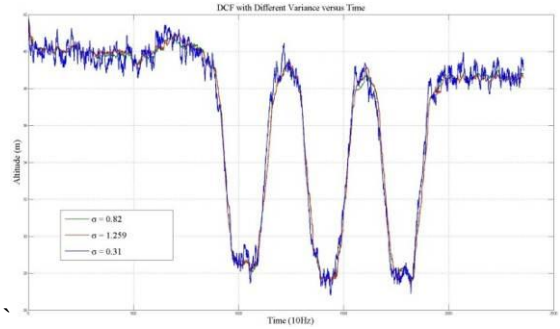


Figure 4. DVF for altimeter with different variance versus time.

For the blue line data, the variance is relatively small and therefore the DVF cannot eliminate the small noise and easily cause data overshoot. For bigger variance, such as red line data, the response will be too slow to reach the actual condition.

III. STEPPER MOTOR TRACKING CONTROL

The ground tracking system is a two-axis rotating mechanism, where the 23 dBi directional antenna is mounted with good alignment to mechanism axis, as shown in Figure 5. Limited by the GPS data update rate, the tracking algorithm has to be capable for estimating the position at the airborne terminal using the moving rate to achieve adjustable control frequency. The microcontroller outputs a digital signal, which is rapid enough to drive rotors to rotate to the desired angle displacement.

Figure 5. Two-axis Stepper Motor Rotation Platform.

 ISBN: 978-1-61208-305-6

However, as the rotor might be in rest state, the starting inertia of each axis will make the mechanism in sudden shock and cause damage to the mechanism. Besides, as the gear backlash increases, the tracking effect will make observable error in the perfect alignment. G. Hilton et al. [13] did the comparison about the terminal's Receiving Sgnal Strength (RSS) change due to different tilt and rotation angle of antenna. The result indicate that as the terminal's antenna polirization is match to transmitter side, terminal can get the maximum RSS. Therefore, the control strategy of two-axis platform become the main issue to avoid the antenna misaligment, which cause high BER in microwave link [14,15]. In order to control the platform smoothly, an appropriate signal output frequency has to be carefully designed. In this paper, a control loop to allow adjustable signal output interval and the unit rotation step for rotors are designed into the tracking control system for MAB.

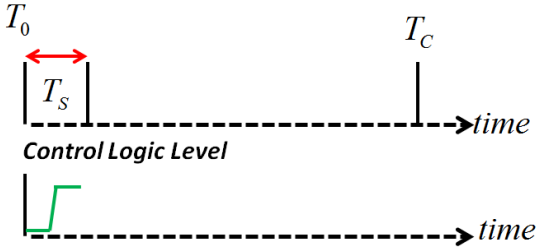


Figure 6. Control sequence.

Figure 6 represents the control sequence, where T_c is the control loop interval and T_s is the control signal output period. For each time after calculation, 2-axis mechanism has the target rotation displacement θ_d, so the adjusted unit step θ_c can be represented as:

$$\theta_c = (\frac{\theta_d}{T_C})T_S \quad (11)$$

With appropriate calculation for each time control core to drive the platform, the adjustable unit step can make each control loop well engaged with less mechanical shaking by decreasing the times of acceleration and deceleration.

However, the adjustable unit step control can smooth the mechanism rotation only. When the airborne terminal has the fast movement in azimuth angle, the rate controller inside control core should be added to avoid the large phase delay of tracking result. The control function can be accomplished by changing the control interval T_c. It should be the function of moving rate, distance and the heading of the airborne terminal. The proposed concept of adaptive control interval is based on that the airborne terminal is moving on ULA or UAV. Its heading information can be used to determine the direction, like North or East, to change faster in the next instants. For example, in azimuth angle tracking, if the heading in T_0 is ψ_1, then, the velocity of the airborne terminal is V_1, the mechanism tracking angle is θ_T. By trigonometric function, the velocity vector of the radial and tangent direction of the airborne terminal with respect to the ground terminal can be estimated as:

$$V_t = V_1 \cos(90 - \theta_T + \psi_1) \quad (12)$$

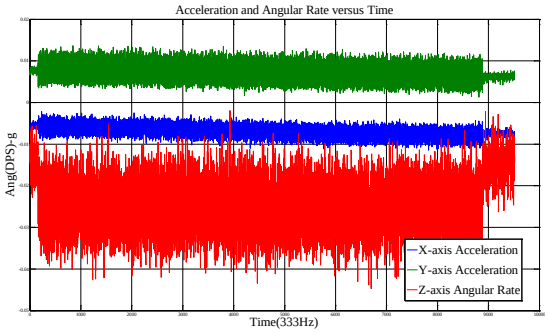
$$V_r = V_1 \sin(90 - \theta_T + \psi_1) \quad (13)$$

Therefore, from (12) and (13), the tracking unit rotation speed in azimuth angle has positive relation with the tangent speed of the airborne terminal. Meanwhile, as its speed increases, the control interval should be decrease to avoid tracking delay. The relation of T_c and tuning gain K_a can be represented as below:

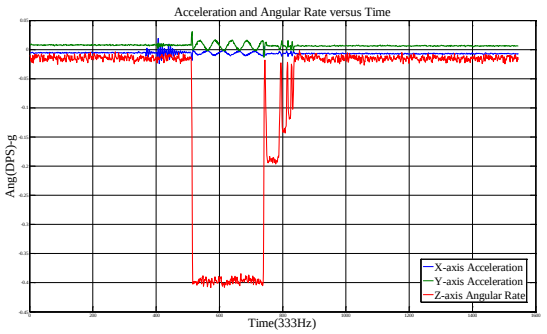
$$T_C(\psi, V, \theta_T) = T_{C0} - \left| \frac{K_a}{V_t} \right| \quad (14)$$

T_{c0} is the default control interval and the optimized motor controller will be the combination of adjustable unit step and rate controller.

In the motor controller tests, the angle displacements are setting at 50.72°, 15.84°, and 0.86°, separately. The figures show the difference of using motor controller or not. The MCU sends 1000Hz control signal to drive the motor directly at unit step equal to 0.02°. Without motor controller, the vibration is large as Figure 7(a). While with motor controller, the vibration appears much reduced in Figure 7(b). In Figure 7(b), when the controller changes the motor's speed, the response drops abruptly, and returns to convergence shortly. Their scales refer to per unit of response. Likewise, Figuress 8 and 9 shows the control results by different angle displacements with changing speed.

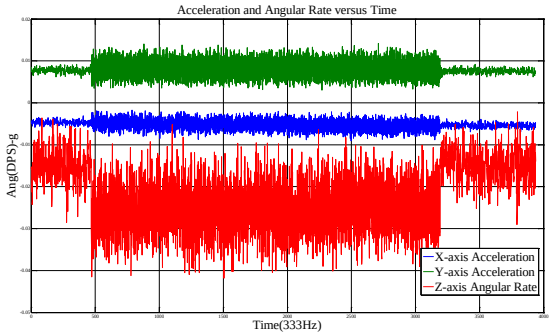


(a) Without controller

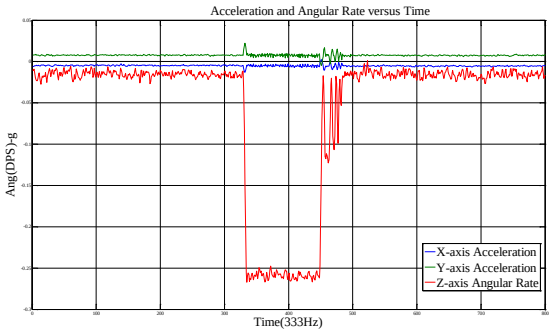


(b) With controller

Figure 7. 50.72° displacement without and with motor controller.

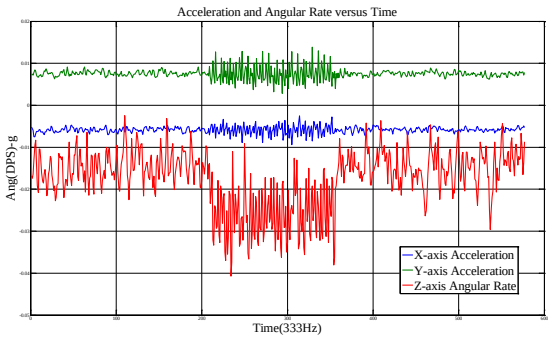


(a) Without controller

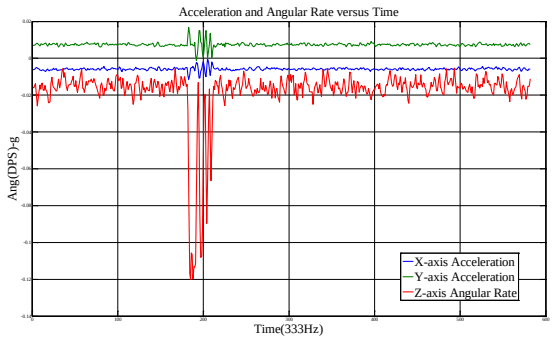


(b) With controller

Figure 8. 15.84˚ displacement without and with motor controller.



(a) Without controller



(b) With controller

Figure 9. 0.86˚ displacement without and with motor controller.

Comparing these three results, as the tracking system without motor controller, the mechanism shaking occurs continuously while the motor is rotating. The time constant of the whole rotation is longer than that with motor controller, especially when the rotation displacement is large.

IV. FLIGHT TEST VERIFICATION

The verification microwave module is installed on unmanned ULA JJ2710, on its wing top, as shown in Figure 10. Two 12 dBi omni-directional antennas were installed perpendicularly to fit the MIMO system. The ground terminal tracking unit was implemented as shown in Figure 11, both azimuth and elevation axis are well balanced to reduce the redundant loading of stepper motor.

Figure 10. Airborne terminal antenna on unmanned ULA JJ2710 for test.

Figure 11. Ground terminal tracking unit with microwave module.

The ground terminal tracking system is calibrated with laser to assure that the rotation original point has no offset, as shown in Figure 12.

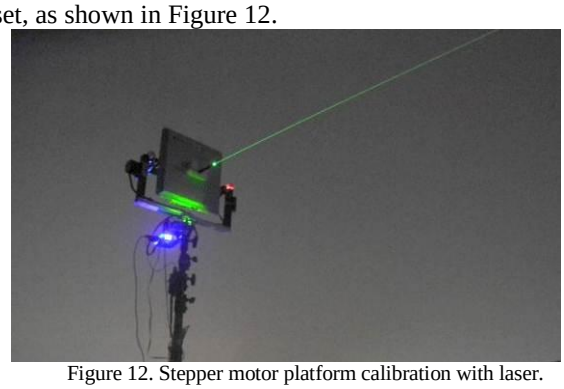


Figure 12. Stepper motor platform calibration with laser.

Flight path was designed to ensure both side of microwave module antenna pattern can be overlapped. Therefore, the ground terminal is located at the mountain side at altitude 468 m (MSL) in southern Taiwan. In this

 ISBN: 978-1-61208-305-6

resting phase, the ULA is flying to approximate the same altitude to track the ground terminal. To verify that the tracking system is capable to maintain constant microwave link, the fight paths and distances are predetermined to check the microwave link. The test distances are 2.5 km and 6 km, respectively, as shown with flight path in Figure 13.

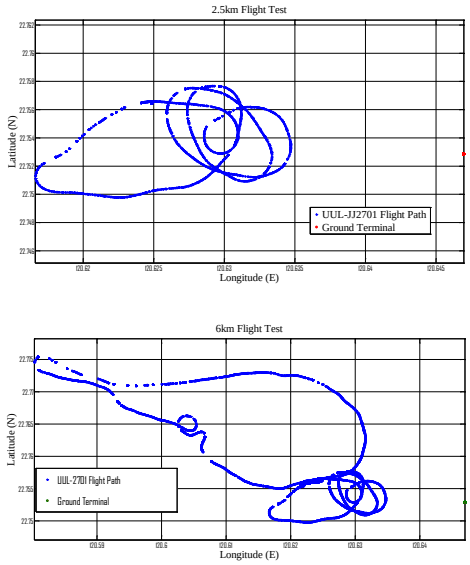
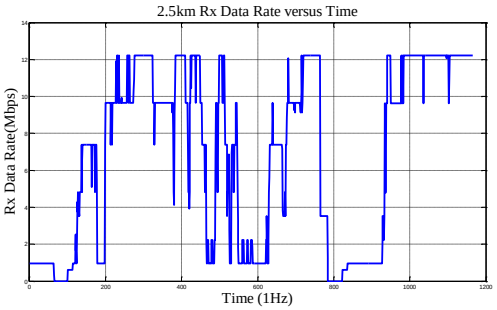
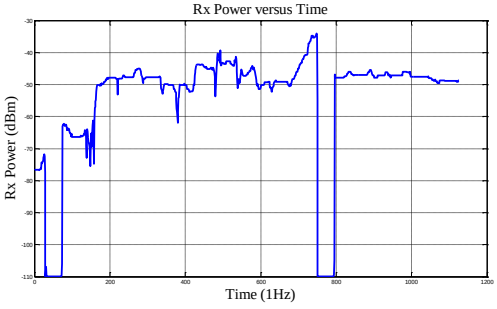


Figure 13. Flight test paths from 2.5 km and 6 km.

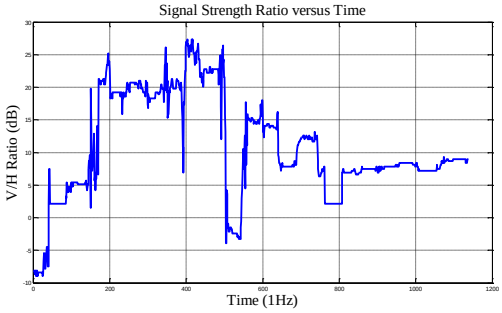
The preliminary flight test results are shown in Figures 14 to 17, for tracking mechanism without or with optimized motor controller. The tracking steps are threefold by take-off, searching, and link up. The tracking mechanism is capable for auto-interval tuning and unit step changing function to precisely track the airborne target.



(a) Rx Data Rate

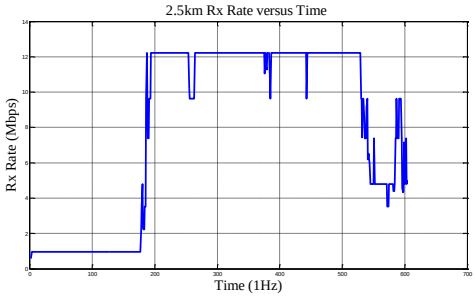


(b) Rx Power



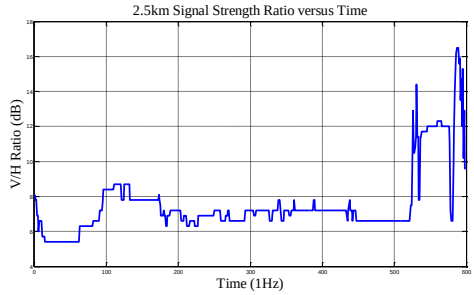
(c) Signal Strength Ratio

Figure 14. Tracking results to 2.5 km without optimized motor controller



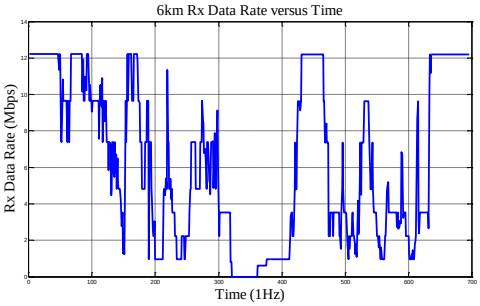
(a) Rx Data Rate

(b) Rx Power

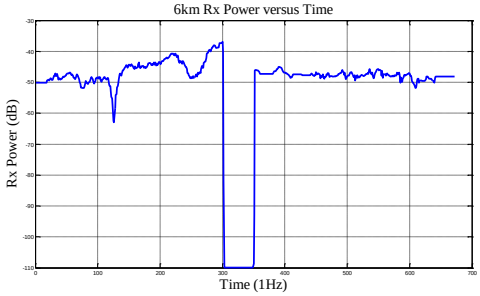


(c) Signal Strength Ratio

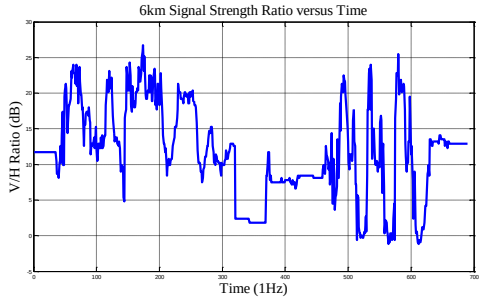
Figure 15. Tracking results to 2.5 km with optimized motor controller



(a) Rx Data Rate



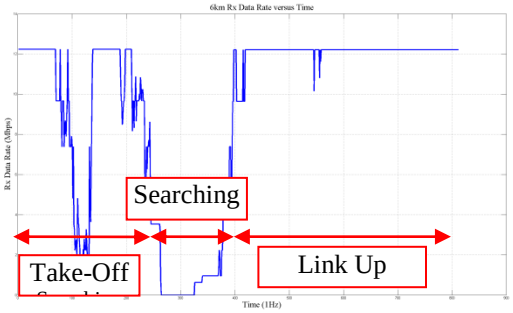
(b) Rx Power



(c) Signal Strength Ratio

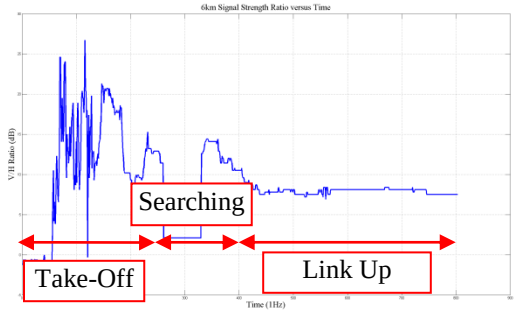
Figure 16. Tracking results to 6 km without optimized motor controller

Figure 17 marks the connection steps from ULA take-off to cruise. During take-off to searching, communication link cannot be made until the UAV turns into stable cruise maneuvering.



(a) Rx Data Rate

(b) Rx Power



(c) Signal Strength Ratio

Figure 17. Tracking results to 6 km with optimized motor controller.

It can be noticed that for longer distance between the airborne terminal and the ground terminal, the overall tracking performance appears better than the closer tests. As the controller core has optimized control algorithm, not only the data rate becomes relatively steady but also the signal strength ratio gets much improved. It is evident that the tracking result is more accurate. The signal strength ratios from both terminals present that the vertical and horizontal polarization antennas are aiming correctly to each other.

V. CONCLUSION

In this paper, a precision antenna alignment tracking system was designed and implemented to Sky Net Project. By considering Earth curvature, a corrective target tracking algorithm is formulated for the vertical and horizontal polarization antennas. Flight tests are carried with a prescheduled flight path to establish microwave link from the airborne terminal to the ground terminal. With optimized dual-axis motor controller, the tracking antenna alignment is proven with feasible results. The alignment problem for microwave transmission can be overcome by the proposed two-axis tracking unit with DVF to filter out the sensor's white noise and unpredicted data error. The test result shows that the tracking unit and antenna implementation can be used to provide the microwave application. It is found of high possibility to establish microwave air-bridging from airborne for mobile communication relaying.

 ISBN: 978-1-61208-305-6

ACKNOWLEDGMENT

This work is supported from National Science Council for Sky-Net Development under contract NSC-101-2218-E-006-002.

REFERENCES

[1] P. Zhan, K. Yu, and A. L. Swindlehurst, "Wireless Relay Communications with Unmanned Aerial Vehicles: Performance and Optimization ", IEEE Transactions on Aerospace and Electronic Systems, July 4, 2011, pp. 2068-2085.

[2] T. C. Tozer and D. Grace, "High-altitude platforms for wireless communications", Electronics & Communication Engineering Journal, Vol. 13, pp. 127-137, ISBN:0954-0695.

[3] O. Cetin, I. Zagli, and G. Yilmaz, "Establishing Obstacle and Collision Free Communication Relay for UAVs with Artificial Potential Fields", Journal of Intelligent & Robotic Systems, Vol. 69, Issue 1-4, 2013, pp. 361-372

[4] O. Cetin and I. Zagli, "Continuous airborne communication relay approach using unmanned aerial vehicles", J. Intell. Robot. Syst. **65**(1–4), 2012, pp. 549–562.

[5] T. Wypych, R. Angelo, and F. Kuester, "AirGSM: An Unmanned, flying GSM cellular base station for flexible field communications", IEEE Aerospace Conference, March 3-10, 2012, pp. 1-9.

[6] F. J. Rowell, D. Sykes, L. Grieveson, B. Theaker, L. Sundar, and R. H. Cumming, "A Near Real-time System for Continuously Monitoring Airborne Subtilisin-Type Enzymes in the Industrial Atmosphere", Journal of Environment Monitor, September 2007, pp. 33–43.

[7] Taiwan Datum TWD 97: http://wiki.osgeo.org/wiki/Taiwan_datums. retrieved: June, 2013.

[8] World Geodetic System WGS 84: https://en.wikipedia.org/wiki/World_Geodetic_System. retrieved: June, 2013.

[9] P. Coote, R. Mahony, K. Jonghyuk, and T. Hamel, "A Complementary Filter for Attitude Estimation of a Fixed-wing UAV", Intelligent Robots Systems International Conference, September 22-26, 2008, pp. 340-345.

[10] Y. Xiaopin and E. Bachman, "Adaptive-Gain Complementary Filter of Inertial and Magnetic Data for Orientation Estimation", Robotics and Automation (ICRA) IEEE International Conference, May 9-13, 2011, pp. 1916-1922.

[11] W. L. Li, C. Y. Sheng , J. W. Hsu, and C. S. Chen, "Motion and Attitude Estimation using Inertial Measurements with Complementary Filter", the 8th Asia Control Conference (ASCC), May 15-18, 2011, pp. 863-868.

[12] C. Silvestre, P. Oliveira, P. Batista, and B. Cardeira, "Discrete time-varying attitude complementary filter", American Control Conference, June 10-12, 2009, pp. 4056-4061.

[13] G. Hilton, E. Mellios, D. Kong, D. Halls, and A. Nix, "Evaluating the Effect of Antenna Tilt and Rotation on Antenna Performance in an Indoor Environment", 2011 Loughborough Antenna and Propagation Conference (LAPC), 14-15 November, 2011, pp. 1-5.

[14] A. K. Hassan, A. Hoque, and A. Moldsvor, "Automatic Micro-Wave (MW) Antenna Alignment of Base Transceiver Stations", Australasian Telecommunication Networks and Applications Conference (ATNAC), November 9-11, 2011, pp. 1-5.

[15] H. Lehpamer, "How to Build a Reliable and Cost-effective Microwave Network", ENTELEC, Houston, Texas, 2006.

 ISBN: 978-1-61208-305-6

A New Design of Dual Band Fractal Antenna for LEO Applications

Lahcene Hadj Abderrahmane
Centre de Développement des Satellites, Bir El Djir
Algerian Space Agency (ASAL)
Oran, Algeria
e-mail: hadjabderrahmanel@yahoo.fr

Ali Brahimi
Faculté de genie électrique
University of Science and Technology, B.P 1505 USTO
Oran, Algeria
e-mail: brahimiali100@gmail.com

***Abstract*—In this paper, a new design of dual band printed antenna based on Minkowski fractal geometry has been presented. The antenna offers a very light weight, low profile, and very low coast, which satisfy the requirement of Low Earth Orbit (LEO) applications. Some techniques have been used to qualify the printed antenna for space applications. Simulation results show the advantage of Minkowski fractal geometry in terms of multiband and bandwidth enhancement. The proposed antenna operates in S and L, Ultra High Frequency (UHF) band efficiently, and has a small size, so it is useful for small satellite communication applications.**

Keywords- printed antenna;dua lband;LEO;Minkowski fractal geometry; multiband.

I. INTRODUCTION

The modern space industry is focused on the small satellites manufacturing to reduce the cost of the mission. With this in mind, research engineers are concentrated on minimizing the mass and the sub-systems number on board satellite. The radio frequency subsystem requires the development of small size, low cost, lightweight, and compact antennas. Printed antennas are the best candidates to meet these requirements [1-5].

After using the high modes of resonance, printed antennas have the ability to operate in two or more bands simultaneously with very similar performance. The advantage of this option is to minimize the total number of antennas on board satellite. Due to the self-similarity nature of their geometry, fractal is used to design the printed antennas to obtain the multi-band property [6].

For space applications, the antenna must be very reliable, mechanically robust, and able of supporting both random vibration and shock at the launch. In orbit, the antenna must be able to survive in the harsh radiation environment, such as ionizing radiation, cosmic rays, and solar energetic particles. Therefore, the materials of the antennas manufacturing must be carefully chosen [7].

As applications, the European Student Earth Orbiter (ESEO) satellite communicates at 2.080 GHz, 2.260 GHz, and bears a total of six microstrip antennas for command and telemetry.

S-band patch antennas are used for communication by the commercial Surrey Satellite Technology Limited (SSTL) micro-satellite, the antenna has a 4.9 dBi gain, main lobe beamwidth equal to 80°, and it has been used for command uplink [8].

A dual-polarized broadband antenna array is presented in [9]. The operative bandwidth is from 3.15 to 3.25GHz, and the peak measured gain is approximately 19 dBi. The proposed antenna has potential applications in Synthetic Aperture Radar (SAR), remote sensing, and wireless communications.

The goal of this work is to use the Minkowski fractal geometry [21, 22]. to design a dual band antenna for LEO applications.

In Section 2, we formulate the concept of fractal geometry to design the proposed antenna. In Section 3, we describe the geometry of the dual band printed antenna, principle, and procedure. Finally, simulation results are shown in Section 4.

II. FRACTAL GEOMETRY

Fractals are used in several areas: statistical analysis, modelling nature, coding, and compression. As they can affect fine structures, fractals have found a good place in art and architecture. In the last two decades, researchers have used fractals in the field of electromagnetism, especially in the antenna design [10].

Fractal, meaning broken or irregular fragments, was originally used by Mandelbrot to describe a family of complex shapes that possess an inherent self-similarity or self-affinity in their geometrical structure.

Generally, using fractal geometries in antennas tends to miniaturize their physical sizes and produce multiband response [11-13].

The first development of the antenna based on fractal geometry has been introduced by Cohen [14], who later founded the company Fractal Antennas Inc. Cohen tried to exploit the usefulness of different pre-fractal geometries empirically, namely Koch curves [10], the curve of Minkowski and Sierpinski carpet [21-22].

 ISBN: 978-1-61208-305-6

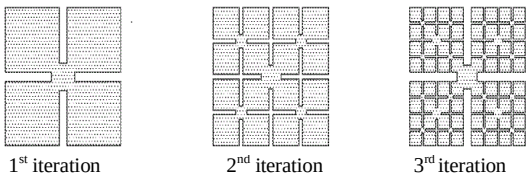


Figure 1. Generation of Minkowski fractal geometry.

In [15], a new microstrip fractal antenna for high impedance matching and bandwidth has been suggested. To improve the antenna performance, many authors proposed using the Defected Ground Structure (DGS) on microstrip antenna [16- 20].

A. *The geometry of Minkowski*

The process of generating fractal geometry is simple, starting with an initial geometric shape called 'initiator' or 'generator', the process is iterative. As a first iteration, each part of the initiator is replaced by a reduced form of the initiator, that is to say, one proceeds to a decrease of scale [21].

For the fractal geometry of Minkowski, from a square, a rectangle of dimension w1 (slot width) × w2 (indentation width) is cut down from the middle of the edge of each side of the square, the generation process of Minkowski pre-fractal geometry is shown in Fig. 1.

Therefore, the circumference (p), increasing with the number of iterations, is given by [23]:

$$p_n = (1 + 2a_2).p_{n-1} \quad (1)$$

$$a_1 = w_1/L_0$$

With; $a_2 = w_2/L_0$; $2(0.5(1-a_1))^D + 2a_2^D + a_1^D = 1$

p_n: the circumference according to the order of iteration.
a_1: *side ratio.*
a_2: *aspect ratio.*
w_1: *slot width.*
w_2: *indentation width.*
L_0: *the length of the side.*
D: the fractal dimension

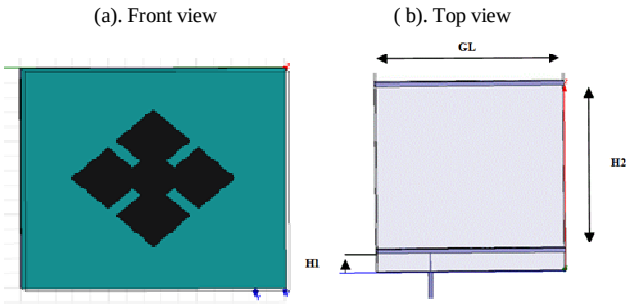


Figure 2. HFSS model of the 1st iteration Minkowski pre-fractal antenna

TABLE I. PHYSIC CHARACTERISTIC OF ROGER RT/DOUROID 5870

Variable	ε_r	tan(δ)	TML (%)	CVCM (%)
Roger RT/douroid 5870	2.33	0.012	0.05	0

B. *Sizing:*

Sizing of the rectangular patch antennas are based on the transmission line model. For the circular patch antennas, the sizing is based on the cavity model. To obtain the antenna dimension values (L,W) of rectangular and circular patch antennas, we use MATLAB for the simulation purposes. The program input are the physical characteristics and material values of the antenna, the dielectric substrate, the thickness of the substrate in mm, the conductivity of the radiating element metal, the loss tangent of the dielectric substrate, and also, the operating frequency [22].

III. ANTENNA DESCRIPTION

For the proposed models, we used the substrate Roger RT / douroid 5870, permittivity ε_r = 2.33, and insertion loss tan (δ) = 0.012. The physical characteristics of this material are shown in Table 1. These characteristics indicate that the Roger RT / douroid 5870 can be used in the spatial domain, the Total Mass Loss (TML) is equal to 0.05% (less than 1%), and Collected Volatile Condensable Materials (CVCM) is 0% (lower than 0.1%). These values satisfy the National Aeronautics and Space Administration (NASA) requirements for the use of materials in space [14].

The geometry of the proposed antenna is shown in Fig. 2 (the simulation model of the first iteration Minkowski pre-fractal antenna), where a diamond patch of length L_0 = 32.95 mm, is placed coplanar with a finite ground plane, which has a square shape of length G_L=75mm.

To obtain circular polarization, the patch was fed by two microstrip lines with orthogonal phase shift of 900. The microstrip lines are printed on the substrate and connected by a 50 Ω coaxial cable. The dielectric substrate used is type Roger RT / duroid with relative permittivity ε_r =2.33 and thickness t = 1.6 mm.

To increase the gain, a parasitic element is used as a patch director; also, upper layer substrate similar to the primary layer was created to the parasite patch [23].

In order to achieve the spreading of the bandwidth, the thickness of the substrates is increased by putting an air layers (ε_r = 1), one between the ground plane, the substrate, the other between the substrate and the substrate whose heights H1=7.1mm and H2 = 63mm, respectively.

The technique of dual power supply by microstrip will be used in order to obtain the circular polarization.

 ISBN: 978-1-61208-305-6

IV. SIMULATION RESULTS

The particular geometry of the fractal antenna and electromagnetic characteristics gives self-similarity that may be used for obtaining the multiband fractal antenna. Due to their geometric complexity, it is very difficult to predict the required performance by using numerical methods. All these methods are based on solving discrete forms of Maxwell field equations.

In this work, we opted for the simulator Ansoft High Frequency Structure Simulator (HFSS) 13.0. The technique used by the software is based on the finite element method. Ansoft HFSS can be used to calculate parameters, such as S-Parameters, Resonant Frequency and Fields. Apart from the normal view design, it provides a 3D view, which is an added advantage.

To investigate the effect of the fractal geometry on the multiplicity of bands, we represent in each case:

- The return loss by taking the maximum value equal to -10 dB.
- The Voltage Standing Wave Ratio (VSWR) to determine the operating frequency taking the maximum value of less than 2 dB.

Fig. 3 shows the variations of the return loss (S11) and the VSWR of the first iteration Minkowski fractal antenna versus frequency. It is noted that the antenna operates in two different frequencies. For VSWR<2, the first frequency is f1=2.0602 GHz corresponds to VSWR1=0.4397, and the second frequency is f2=3.4135 GHz corresponds to VSWR2=0.323.

The radiation pattern of the first iteration Minkowski fractal antenna, for the two angles phi=0° and phi=90°, is shown in Fig. 4. It is observed that, for the two plans of the electric field E (phi=0°) and the magnetic field H (phi=90°), the antenna aperture is about 60°, which gives a quasi-hemispherical radiation pattern.

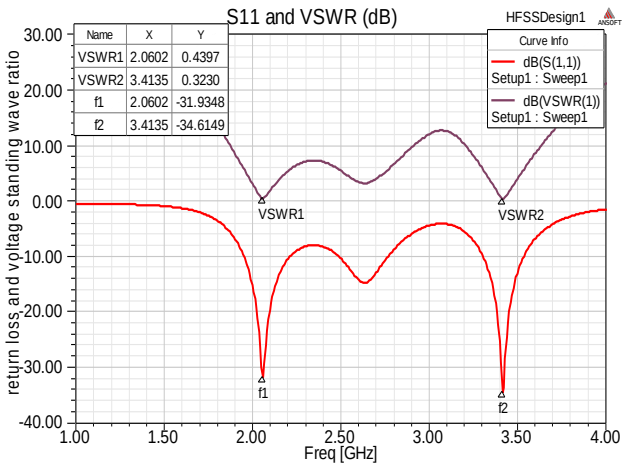


Figure 3. Variations of return loss and voltage standing wave ratio of the first iteration Minkowski fractal antenna

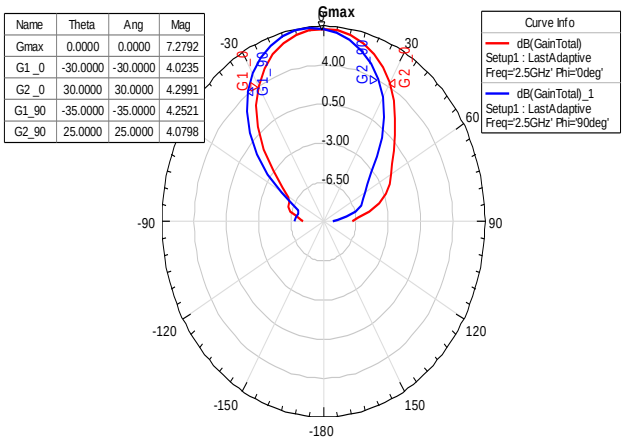


Figure 4. Radiation pattern of the first iteration Minkowski pre-fractal antenna.

Fig. 5 shows the 3D gain variation pattern of the first iteration Minkowski antenna. It is noticed that the maximum gain of the antenna is Gmax=7.357dB in the Z axis direction (phi=0°).

The return loss (S11) of the two Minkowski antennas for the first and the second iteration are presented in Fig. 6. For | S11 | < -10 dB, the second iteration Minkowski antenna operates for four frequencies in two bands: L band for the frequency f_1 = 1.8722 GHz: | S11 | = -20 dB, and S band for f_2 = 2.2632 GHz: | S11 | = -13.36 dB, f_3 = 2.4962 GHz: | S11 | = -24.35 dB, and f_4=2.9323 GHz: | S11 | = -20.26 dB.

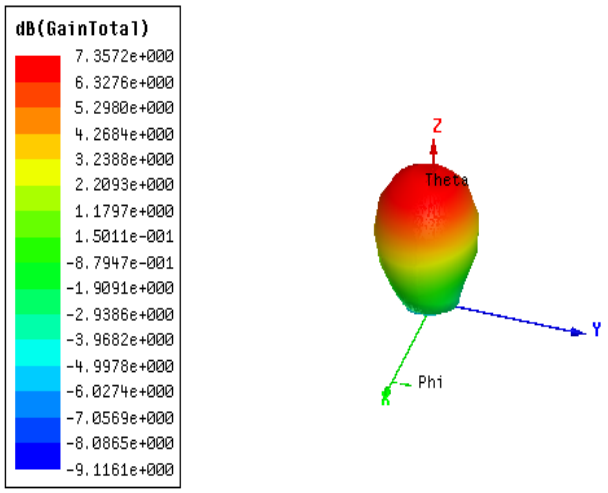


Figure 5. 3D gain variation pattern.

 ISBN: 978-1-61208-305-6

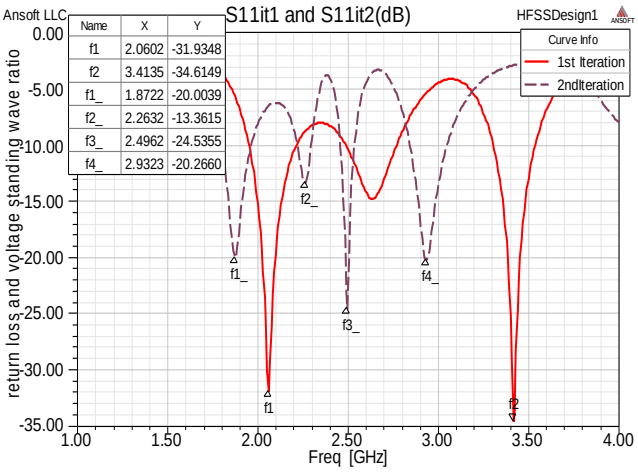


Figure 6. S11 of the first and the second iteration Minkowski fractal antennas

By comparison with the first iteration Minkowski antenna, we notice the appearance of two new frequencies.

For the condition VSWR<2, the frequency f2 = 2.2632 GHz (VSWR2=3.78) does not present a resonance frequency, we can thus conclude that the second iteration Minkowski antenna presents three resonances frequencies (Fig. 7).

Fig. 8 represents the model of the third iteration Minkowski fractal antenna using HFSS (the relative axis is tilted by 45° compared to the principal axis).

However, the gain decreases when the iteration number increases. This is justified by the increasing of the antenna input impedance, which leads to antenna and microstrip line mismatch.

The radiation pattern of three iterations for phi=90° and phi=0°, is shown in Fig. 9 and Fig. 10, respectively.

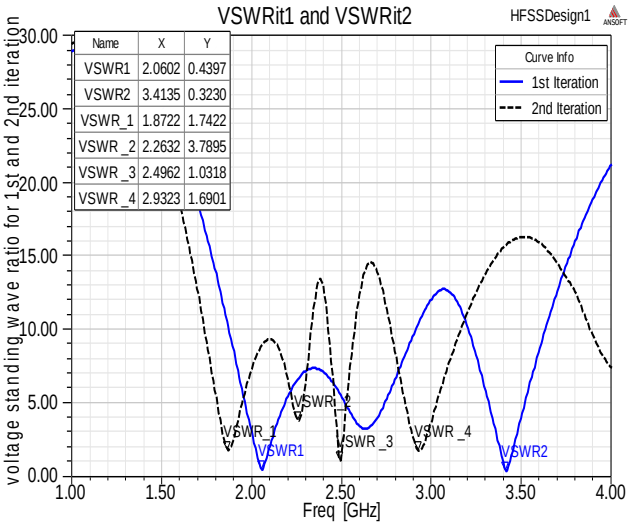


Figure 7. VSWR of the first and the second iteration Minkowski fractal antennas

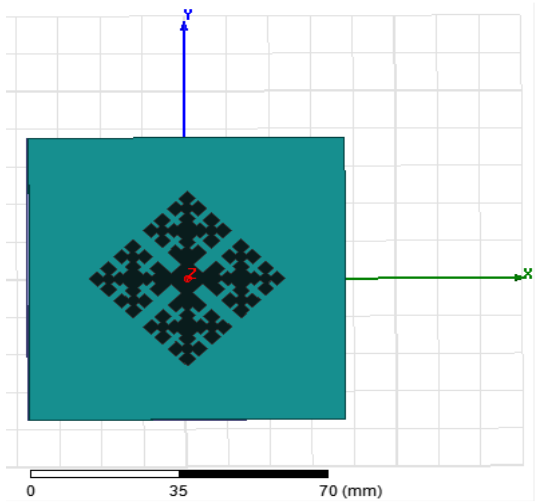


Figure 8. HFSS model of the third iteration Minkowski fractal antenna

According to Fig. 10, it is observed that the antenna aperture increases relatively with the iteration number; this shows the advantage of using the fractal geometry.

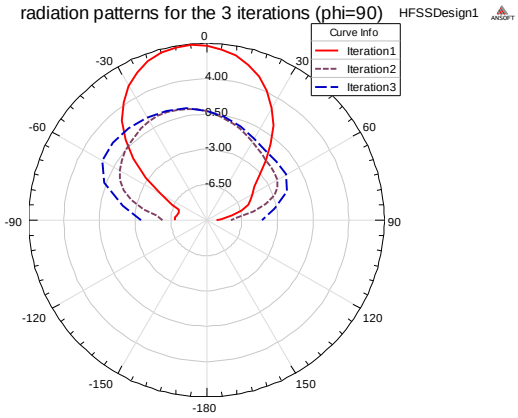


Figure 9. The radiation pattern of three iterations for phi=90°

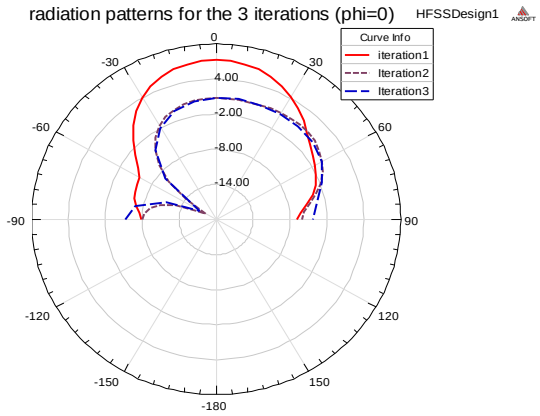


Figure 10. The radiation pattern of three iterations for phi=0°

V. CONCLUSION AND FUTURE WORK

In this paper, a new structure of printed antenna based on the Minkowski fractal geometry has been presented. The suggested structure is made up of two layers of substrate and a parasitic element of patch in order to increase the gain. A separation by layers of airs is done to increase the antenna bandwidth.

A fractal geometry study effect on the printed antennas is done; a comparative study was drawn up to conclude that the use of fractal geometry has several advantages, such as the multiplicity of band, and the increase in the antenna bandwidth.

The proposed antenna is characterized by reduced size, low cost, low profile, and rigid structure.

The third iteration Minkowski pre-fractal antenna operates in S and L bands, presents a moderate gain, and a quasi-hemispherical radiation pattern. This antenna can be used in telemetry, tracking, and control for satellite Earth observation.

Finally, we may conclude that the space parameters and structural of the antenna are very affected on the RF performance of the printed antenna. The advantage of the fractal geometry is the property multi-band, but this is limited by the structural performance of the antenna, and then by the form of the radiating element.

As future work, we can extend this work to VHF/UHF bands, a large number of potential applications arise. The low input resistance for the antenna using fractal geometry can be improved by feeding the antenna suitably. The suggested antenna presented in this work needs an optimization algorithm for radiation study.

REFERENCES

[1] J. Zhou, B. Liang, B. You, Q. Liu, and X. Yan, “A Fractal Microstrip Array Antenna with Slots Feeding Network for DTV Reception,” PIERS Proceedings, Stockholm, Sweden, 2013, pp. 262-265.

[2] A. Sabban, “Applications of MM wave microstrip antenna arrays," Signals, Systems and Electronics, Vol. 07, 2007, pp.119-122.

[3] M. Shah and M. K. Suaidi, “Design of 1x2, 1x4, and 2x2 dual polarization microstrip array antenna," Proceedings of IEEE 6th National Conference on TelecommunicationTechnologies, 2008, pp. 113-116.

[4] M. H. Awida, “Substrate-integrated waveguide Ku-band cavity backed 2£2 microstrip patch array antenna," Antennas and Wireless Propagation Letters, vol. 8, 2009, pp. 1054-1056.

[5] R. Azadegan, “A Ku-band planar antenna array for mobile satellite TV reception with linear polarization," IEEE Transactions on Antennas and Propagation, vol. 58, no. 6, 2010, pp. 2097-2101.

[6] K. J. Vinoy, “Fractal shaped antenna elements for wide- and multi- band wireless applications,” A PHD thesis in Engineering science and mechanics, Pennsylvania state University, the graduate school College of engineering, August 2002.

[7] S. Gao, M. Brenchley, and M.Unwin, "Antennas for small satellites," 2008 Loughborough Antennas & Propagation Conference, Loughborough, UK, 2008, pp. 66-69.

[8] J. Sosa-Pedroza, F. Martinez-zuñiga, and M. Enciso-Aguilar, Planar Antennas for Satellite Communications, Satellite communications, 2010, pp. 367-394.

[9] R. Di Bari, G. S. Brown, M. Notter, D. Hall, and C. Underwood, "Dual-Polarized Printed S-Band Radar Array Antenna for Spacecraft Applications," Antennas and Wireless Propagation Letters, IEEE, vol. 10, 2011, pp. 987-990.

[10] D. Li and J. Mao, “A Koch-Like Sided Fractal Bow-Tie Dipole Antenna,” IEEE Transactions on Antennas and Propagation, vol. 60, no. 5, May 2012, pp.2242-2251.

[11] L. Lizzi, R. Azaro, G. Oliveri, and A. Massa, “Printed UWB Antenna Operating Over Multiple Mobile Wireless Standards,” IEEE Antennas and Wireless Propagation Letters, vol. 10, 2011, pp. 1429-1432.

[12] M. Jahromi, A. Falahati, and R. Edwards, “ Bandwidth and Impedance Matching Enhancement of Fractal Monopole Antennas Using Compact Grounded Co-planar Waveguide,” IEEE Transactions on Antennas and Propagation, vol. 59, no. 7, July 2011, pp. 2480-2487.

[13] K. Singh, V. Grewal, and R. Saxena, “Fractal Antennas: A Novel Miniaturization Technique for Wireless Communications,” International Journal of Recent Trends in Engineering, vol 2, no. 5, November 2009, pp. 172-176.

[14] N. Cohen, Fractal Antennas: Part 2, Communications Quarterly, Summer, 1996, pp. 53-66.

[15] S. Rani and A. P. Singh, “On the Design and Optimization of New Fractal Antenna Using PSO,” International Journal of Electronics, DOI:10.1080/00207217.2012.743089, Nov.2012, pp. 1383-1397.

[16] J. Verringer and A. Nafalski, “Fractal Antenna Application to Satellite Communications,” International Journal of Applied Electromagnetics and Mechanics, 2001/2002, pp. 271-276.

[17] R. Azaro, L. Debiasi, M. Benedetti, P. Rocca, and A. Massa, “A Hybrid Prefractal Three-Band Antenna for Multistandard mobile Wireless Applications,” IEEE Antennas and Wireless Propagation Letters, vol. 8, 2009, pp. 905-908.

[18] S. Rani and A. P. Singh, “Fractal Antenna with Defected Ground Structure for Telemedicine Applications,” International journal on Communications, Antenna and Propagation, vol. 1, 2012, pp. 1-15.

[19] D. Schlieter and R. Henderson, “High Q Defected Ground Structures in Grounded Coplanar Waveguide,” Electronic Letters, vol. 48, no. 11, May 2012, pp. 635-636.

[20] S. Kakkar and P. S. Rani, “New Antenna with Fractal Shaped DGS for Emergency Management Applications,” International Journal of Advanced Research in Computer Science and Software Engineering vol. 3, no. 3, March 2013, pp. 721-724

[21] H. Xu, G. Wang, X. Yang, and X. Chen, “Compact, Low Return-Loss, and Sharp-Rejection UWB Filter Using Sierpinski Carpet Slot in a Metamaterial Transmission Line, International,” Journal of Applied Electromagnetics and Mechanics, 2011, pp.253-262.

[22] J. K. Ali, “A New Reduced Size Multiband Patch Antenna Structure Based on Minkowski Pre-Fractal Geometry,” Journal of Engineering and Applied Sciences, JEAS, vol. 2, no. 7, 2007, pp. 1120-1124.

[23] L. Jianzhou, G. Steven, and X. Jiadong, "Circularly Polarized High-Gain Printed Antennas for Small Satellite Applications," International Conference on Microwave Technology and Computational Electromagnetics (ICMTCE2009) Beijing, China, , ISBN: 978 1 84919 140 1, Nov. 2009, pp. 76-79

 ISBN: 978-1-61208-305-6

Coexistence of Earth Station of the Fixed-Satellite Service with the Terrestrial Fixed Wireless System in 8 GHz Band

Jong-Min Park, Nam-Ho Jeong, Dae-Sub Oh and Bon-Jun Ku
Satellite & Wireless Convergence Research Department
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
jongmin@etri.re.kr, nhjeong@etri.re.kr, trap@etri.re.kr, bjkoo@etri.re.kr

***Abstract*—This paper presents evaluation results of coexistence between earth station of the fixed-satellite service and terrestrial fixed wireless system in 8 GHz band. The evaluation has been made based on a methodology and system characteristics assumed for analysis on frequency sharing basis. The result could be useful when new frequency allocation to the fixed-satellite service is considered in the frequency band to which the terrestrial fixed service is already allocated.**

Keywords-fixed-satellite; earth station; terrestrial fixed wiereless system; interference; coexistence

I. INTRODUCTION

Fixed-satellite service (FSS) is the official classification for communications using geostationary satellites that provide broadcast feeds to television stations, radio stations and broadcast networks. FSSs also transmit information for telephony, telecommunications and data communications [1].

In order to deploy a satellite network providing various services in a wide area as mentioned above, spectrum and orbit resources are essential. Since they are limited natural resources [2], it is very important to use them efficiently and economically.

The frequency bands 7.25-7.75 GHz and 7.9-8.4 GHz are allocated worldwide to the FSS in the direction of space-to-Earth and Earth-to-space, respectively. These bands or parts of them are also allocated worldwide to other services such as the fixed and mobile services, the meteorological-satellite service and the Earth exploration-satellite service (space-to-Earth). These bands have been generally used for military or satellite imagery. At World Radiocommunication Conference held in 2012 (WRC-12), some countries reported a shortfall of spectrum available for their current and future applications in these bands. The additional bandwidth requirements for data transmission on these next-generation satellites were estimated around a maximum of 100 MHz. To meet the requiremets, it was decided that WRC-15 should consider possible new allocations to the FSS in the frequency bands 7.15-7.25 GHz (space-to-Earth) and 8.4-8.5 GHz (Earth-to-space). When considering any additional possible frequency allocations to any space services, compatibility studies to ensure adequate protection of terrestrial services as in [3].

This paper presents the possibility of coexistence between earth station of the FSS and terrestrial fixed wireless system (FWS) in the band 8.4-8.5 GHz. The evaluation has been made based on a methodology and system characteristics assumed for analysis on frequency sharing basis as presented in Section II. Section III analyzes the coexistence of the FSS earth station with the FWS based on the results of interference calculation. Finally, we provide our conclusion form the study results.

II. METHODOLOGY AND SYSTEM CHARACTERISTICS

A. Interference Scenario and Methodology

Fig.1 shows the interference scenario considered in the study.

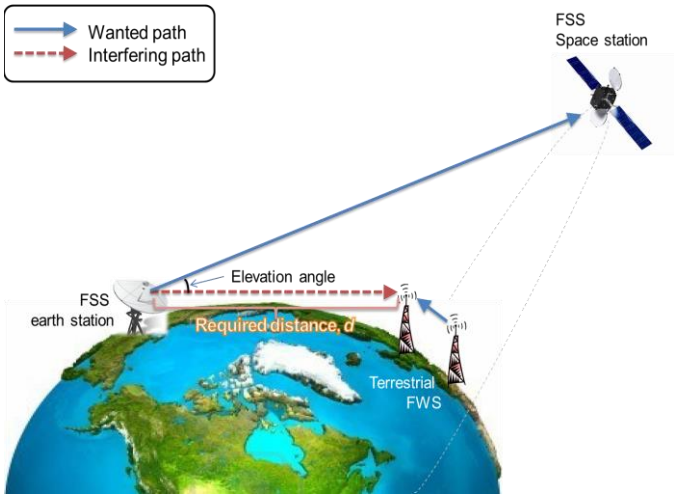


Figure 1. Interference scenario

The received interference power density at the receiver of the FWS is calculated using (1).

$$P_r = P_t + G_t(\theta_t) - L_b(p\%) + G_r(\theta_r) \quad (1)$$

where:

P_t : Transmitter power density of FSS earth station (dBW/MHz);

$G_t(\theta_t)$: Antenna gain of FSS earth station towards FWS system (dBi);

θ_t : Angle between the main emission of FSS earth station and interference reception (degrees);

 ISBN: 978-1-61208-305-6

G_r (θ_r): Antenna gain of FWS towards FSS earth station (dBi);

θ_r : Angle between the main beam of FWS receive antenna and interference source (degrees);

L_b (p%): Path loss level not exceeded for p% time (dB).

In order to evaluate if the received interference power density at the receiver of the FWS from the emission of FSS earth station can meet the protection criterion of the FWS, we calculated L_b (p %) and got the required protection distance between interfering FSS earth station and FWS receiver. The required protection distance is determined based on the propagation losses indicated in the methodology which is given in [4] which is widely used in similar studies as in [5].

If FSS in the Earth-to-space direction is to be introduced into bands already heavily used, aggregate interference impacts on the existing services in the bands should be considered as appropriate. *I/N* values for long-term interference of -6 dB or -10 dB, as appropriate, may be applicable where the risk of simultaneous interference from the stations of the other co-primary allocations is negligible and in other cases, a more stringent criterion may be required to account for aggregate interference from all interfering co-primary services [6].

It is possible to apportion allowable interference in digital FWS to the FS, other services and other emissions respectively as 89 %, 10 % and 1 % of the total interference allowance [7]. Allowing 20 % degradation due to total interference, this means that the allowance for other co-primary services is 2 % of the error performance objectives. If only FSS is considered in the band, the FSS portion would then be 2 % of the error performance objective, leading to an allowable *I/N* of −17 dB. If another or two other co-primary service(s) is/are considered as co-primary service(s) in the band, the FSS portion would be 1 % or 0.67 %, leading to an allowable *I/N* of -20 dB or -21.7 dB.

B. System Characteristics for Interference Analysis

Table I presents system characteristics of FSS earth station assumed for interference analysis in the study. We considered five types of FSS earth station for various applications in the FSS.

We assumed antenna pattern for the FSS earth station as in Fig. 2. The earth stations of Type 1 to Type 4 have the same antenna pattern except for the maximum gain as given by (2) for $D/\lambda \geq 50$, where D is antenna diameter and λ is wavelength [8]. The earth station of Type 5 has a different pattern from the others, since the antenna pattern of Type 4 was extended for $D/\lambda < 50$ as given by (3) [9].

$$
\begin{aligned}
G(\varphi) &= G_{max} - 2.5\times10^{-3}\,(D/\lambda\cdot\varphi)^2 && \text{for } 0°\leq \varphi < \varphi_m \quad (2)\\
&= G_1 && \text{for } \varphi_m \leq \varphi < \varphi_r\\
&= 32 - 25\log\varphi && \text{for } \varphi_r \leq \varphi < 20°\\
&= 52 - 10\log(D/\lambda) - 25\log\varphi && \text{for } 20° \leq \varphi < \varphi_b\\
&= 10 - 10\log(D/\lambda) && \text{for } \varphi_b \leq \varphi \leq 180°
\end{aligned}
$$

TABLE I. SYSTEM CHARACTERISTICS OF FSS EARTH STATION

FSS earth station parameters	Units	Type 1	Type 2	Type 3	Type 4	Type 5
Frequency	GHz	8.45	8.45	8.45	8.45	8.45
Maximum transmit output power	dBW	33.0	33.0	27.8	33.0	30.0
Bandwidth	MHz	50	50	50	2	2
Transmit antenna diameter	m	18.0	11.0	5.0	2.5	1.5
Transmit antenna gain	dBi	62	58	51	45	40
Earth station side lobe attenuation	dB	58	54	47	41	29.3
Transmit antenna height	m	15	15	5	5	5
Transmit loss	dB	2	2	2	2	2
Transmit off-axis e.i.r.p.	dBW	35.0	35.0	29.8	35.0	38.7
Transmit off-axis e.i.r.p. density in 1 MHz bandwidth	dBW/ MHz	18.0	18.0	12.8	32.0	35.7

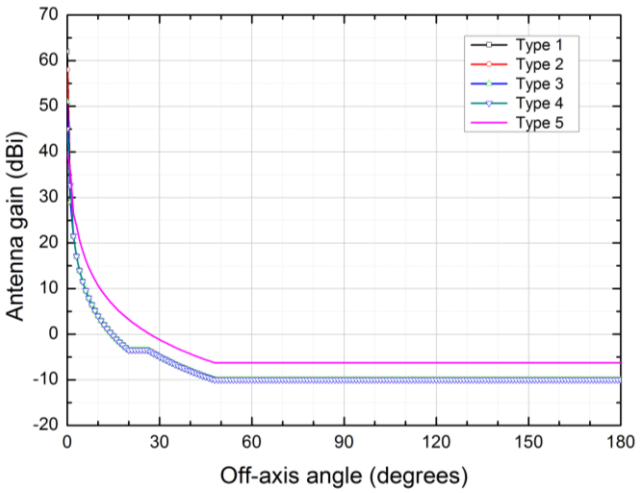


Figure 2. Antenna pattern of FSS earth stations

where:

G_{max} = Maximum antenna gain;

$D/\lambda = (10^{(Gmax/10)} / \eta\,\pi^2)^{0.5}$;

$\varphi_m = 20\,\lambda/D\cdot(G_{max}-G_1)^{0.5}$;

$G_1 = 2 + 15\log(D/\lambda)$; for $D/\lambda \leq 150$

$= -1 + 15\log(D/\lambda)$; for $D/\lambda > 150$

$\varphi_r = 15.85\,(D/\lambda)^{-0.6}$; for $D/\lambda \geq 100$

$= 100\,(\lambda/D)$; for $D/\lambda < 100$

$\varphi_b = 48^{\circ}$.

$$
\begin{aligned}
G(\varphi) &= G_{max} - 2.5\times10^{-3}\,(D/\lambda\cdot\varphi)^2 && \text{for } 0°\leq \varphi < \varphi_m \quad (3)\\
&= G_1 && \text{for } \varphi_m \leq \varphi < \varphi_r\\
&= 52 - 10\log(D/\lambda) - 25\log\varphi && \text{for } \varphi_r \leq \varphi < \varphi_b\\
&= 10 - 10\log(D/\lambda) && \text{for } \varphi_b \leq \varphi \leq 180°
\end{aligned}
$$

Table II presents system characteristics of FWS assumed for interference analysis in the study.

TABLE II. SYSTEM CHARACTERISTICS OF FWS

FWS parameters	Units	Value
FWS receiver antenna gain	dBi	48.6
FWS side lobe attenuation	dB	16
FWS receiver antenna height	m	50
FWS cable loss	dB	3

We assumed antenna pattern for the FWS as given by Fig. 3 using (4) [10].

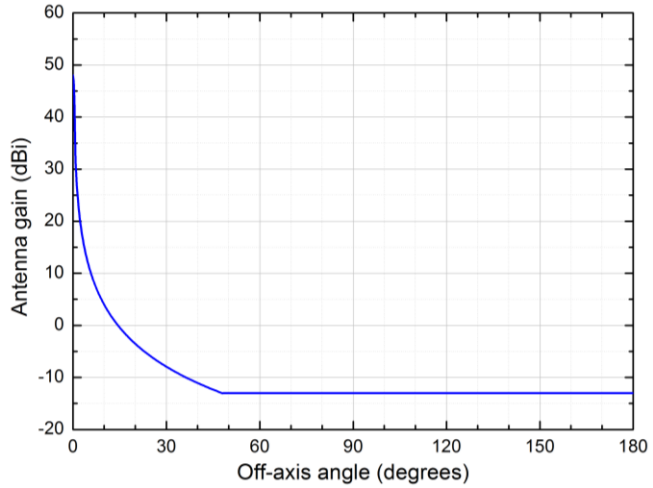


Figure 3. Antenna pattern of FWS

$$G(\varphi) = G_{max} - 2.5\times10^{-3}(D/\lambda\cdot\varphi)^2 \quad for\ 0^\circ < \varphi < \varphi_m \qquad (4)$$
$$= G_1 \quad for\ \varphi_m \le \varphi < \max(\varphi_m, \varphi_r)$$
$$= 29 - 25\ log\ \varphi \quad for\ \max(\varphi_m, \varphi_r) \le \varphi < 48^\circ$$
$$= -13 \quad for\ 48^\circ \le \varphi \le 180^\circ$$

where:

G_{max}: Maximum antenna gain (dBi);

$G(\varphi)$: Gain relative to an isotropic antenna (dBi);

φ: Off-axis angle (degrees);

D: Antenna diameter (m);

λ: Wavelength (m);

G_1: Gain of the first side lobe;

$= 2 + 15 \log (D/\lambda)$;

$$\varphi_m = \frac{20\lambda}{D}\sqrt{G_{max} - G_1} \quad \text{(degrees);}$$

$$\varphi_r = 12.02\,(D/\lambda)^{-0.6} \quad \text{(degrees).}$$

III. CALCULATION RESULTS AND ANALYSIS OF COEXISTENCE OF FSS EARTH STATOIN WITH FWS

We calculated required $L_b(p\ \%)$ to meet the protection criteria of FWS taking into account for propagation model with flat terrain and time percentage, p of 20 % for long-term analysis and finally found the required protection distance creating the required $L_b(20\ \%)$. Fig. 4 shows the calculated $L_b(20\ \%)$ based on the system characteristics provided in the previous section. We assumed the average radio-refractive index lapse-rate through the lowest of the atmosphere ΔN = 45 and the sea-level surface refractivity N_0 = 310. The propagation mechanisms include tropospheric scatter, ducting, fade and gaseous absorption over the path between the location of emission and reception.

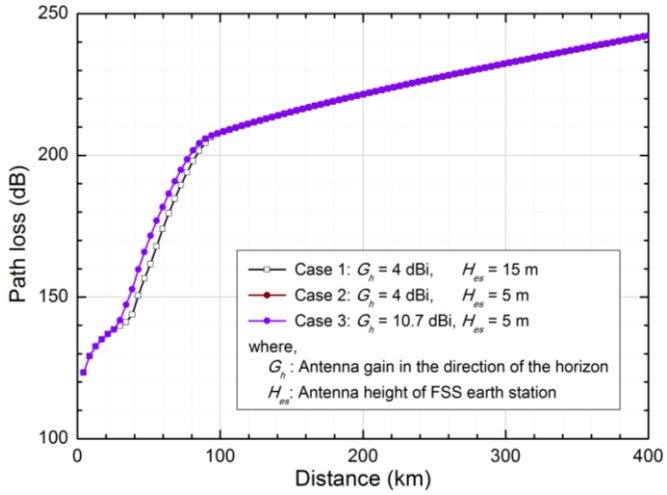


Figure 4. Calculation results of path loss vs. distance

In Fig. 4, Case 1, to which Types 1 and 2 of the FSS earth station belong, shows slightly low path loss compared to Cases 2 and 3 on the distance below 100 km, while Cases 2 and 3, to which Types 3 and 4 and Type 5 belong, respectively, show the same result. It implies that the antenna height of FSS earth station would be a dominant factor for the path loss.

Based on the results given in Fig. 4, we could get the required distance to meet long-term interference level of terrestrial FWS from the emission of FSS earth station. Table III presents the calculation results of the required separation distance.

The results of the static analysis shows that FSS earth station of all types can be compatible with FWS if it would ensure the required protection distances from 79.8 km to 261.5 km.

It should be noted that the calculations were carried out for flat terrain not taking into account the actual path profile of the interfering signal. Since, in real situation, the interference will be additionally decreased due to natural and artificial obstacles, the required distance between FSS earth station and FWS will be less than the calculated results shown in Table III.

 ISBN: 978-1-61208-305-6

TABLE III. CALCULATION RESULTS OF REQUIRED DISTANCE FROM FSS EARTH STATION TO MEET LONG-TERM INTERFERENCE LEVEL OF FWS

Calculation item	Units	Case 1						Case 2						Case 3		
		Earth Station type 1			Earth Station type 2			Earth Station type 3			Earth Station type 4			Earth Station type 5		
Calculated interference power density	dBW/MHz	47.6			47.6			42.4			61.6			65.3		
Allowable I/N	dB	-17	-20	-21.7	-17	-20	-21.7	-17	-20	-21.7	-17	-20	-21.7	-17	-20	-21.7
FWS nominal long term interference criteria	dBW/MHz	-158.5	-161.5	-163.2	-158.5	-161.5	-163.2	-158.5	-161.5	-163.2	-158.5	-161.5	-163.2	-158.5	-161.5	-163.2
Required attenuation	dB	206.1	209.1	210.8	206.1	209.1	210.8	200.9	203.9	205.6	220.1	223.1	224.8	223.8	226.8	228.5
Required protection distance	Km	93.3	107.3	118.2	93.3	107.3	118.2	79.8	84.7	89.0	188.4	214.1	229.1	218.7	245.8	261.5

IV. CONCLUSION

This paper addresses the feasibility of coexistence between transmitting FSS earth station and receiving terrestrial FWS in 8 GHz band. We calculated required separation distance of FSS earth station to meet the long-term interference level of FWS from interference path loss, taking into account interference methodology and system characteristics presented in the previous section.

The results show that the required separation varies up to a few hundreds kilometers if all types of FSS earth station will be deployed. If we apply actual path profile for the calculation, the required distance could be reduced due to natural and artificial obstacles.

Based on the results, we can select a certain type of FSS earth station for coexistence with terrestrial FWS, when we consider new frequency allocation to the FSS.

ACKNOWLEDGMENT

This study was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.

REFERENCES

[1] ITU, Handbook of Satellite Communications, 3rd Edition, 2002.

[2] ITU, Constitution of the International Telecommunication Union, Edition of 2011.

[3] J.M. Park, D.S. Ahn, H.J. Lee and D.C. Park, "Feasibility of Coexistence of Mobile-Satellite Service and Mobile Service in Cofrequency Bands," *ETRI J.*, vol. 32, no. 2, Apr. 2010, pp. 255-264.

[4] ITU, "Prediction procedure for the evaluation of interference between stations on the surface of the Earth at frequencies above about 0.1 GHz," Recommendation ITU-R P.452-14, 2009.

[5] H.S. Jo, "Codebook-Based Precoding for SDMA-OFDMA with Spectrum Sharing," *ETRI J.*, vol. 33, no. 6, Dec. 2011, pp. 831-840.

[6] ITU, "System parameters and considerations in the development of criteria for sharing or compatibility between digital fixed wireless systems in the fixed service and systems in other services and other sources of interference," Recommendation ITU-R F.758-5, 2012.

[7] ITU, "Maximum allowable error performance and availability degradations to digital fixed wireless systems arising from radio interference from emissions and radiations from other sources," Recommendation ITU-R F.1094-2, 2007.

[8] ITU, "Radiation diagrams for use as design objectives for antennas of earth stations operating with geostationary satellites," Recommendation ITU-R S.580-6, 2003.

[9] ITU, Appendix 8 of the Radio Regulations, Edition of 2001.

[10] ITU, "Mathematical model of average and related radiation patterns for line-of-sight point-to-point fixed wireless system antennas for use in certain coordination studies and interference assessment in the frequency range from 1 GHz to about 70 GHz," Recommendation ITU-R F.1245-5, 2012.

 ISBN: 978-1-61208-305-6

A Novel Unambiguous BOC Acquisition Scheme for Global Navigation Satellite Systems

Youngseok Lee and Seokho Yoon*
College of Information and Communication Engineering
Sungkyunkwan University
Suwon, Korea
e-mail: {fortrtwo and *syoon}@skku.edu

***Abstract*—This paper addresses the problem of ambiguity in the acquisition of binary offset carrier (BOC) signals employed in global navigation satellite systems, which is caused by the multiple side-peaks of the BOC autocorrelation function. We first observe that the side-peaks arise due to the fact that the BOC autocorrelation is made up of the sum of the sub-correlations shaped irregularly, and then, propose a novel unambiguous BOC acquisition scheme based on a recombination of the sub-correlations. The proposed scheme is demonstrated to remove the side-peaks completely for any type of BOC signals and to provide a performance improvement over the conventional scheme in terms of the receiver operating characteristic and mean acquisition time.**

Keywords-Acquisition; ambiguity problem; binary offset carrier

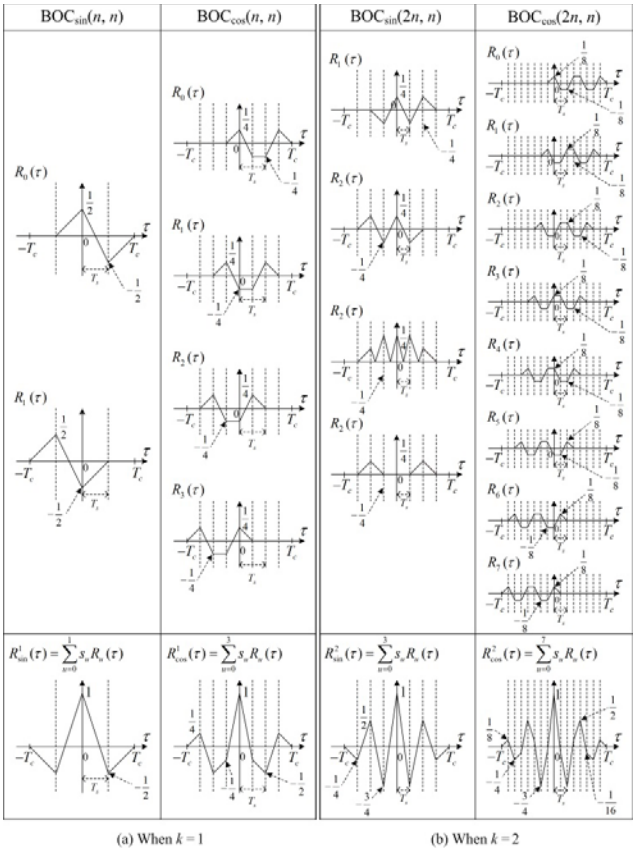
I. INTRODUCTION

Recently, new global navigation satellite systems (GNSSs) such as Galileo and global positioning system (GPS) modernization are being developed to satisfy the increasing demand for GNSS-based services such as location-based service (LBS) and emergency rescue service (ERS) and complement the existing GNSSs such as GPS [1]-[3]. Currently, new GNSSs are designed to use the same frequency band of the existing GNSSs: for example, the E1 and E5 bands of Galileo are overlapped with the L1 and L5 bands of GPS, respectively [1], [4]. Thus, if a Galileo signal is modulated by a conventional scheme such as phase shift keying (PSK) used in GPS, it would suffer from co-channel interference. To overcome these problems, binary offset carrier (BOC) modulation has been proposed, where a high degree of spectral separation between the BOC-modulated signals and the others is achieved by shifting the signal energy from the band center [5]. The BOC signal is generated through the product of a spreading pseudo random noise (PRN) code and a sine-phased or cosine-phased square wave sub-carrier, and denoted by $\mathrm{BOC}_{\sin}(kn,n)$ or $\mathrm{BOC}_{\cos}(kn,n)$ depending on which of the sine-phased and cosine-phased sub-carriers are used, where k and n are the ratios of the PRN code chip period to the sub-carrier period and the PRN code chip rate to 1.023 MHz, respectively [4], [6]. For larger values of k, more separated spectrums are obtained, reducing the co-channel interference more effectively. However, the BOC signal has multiple side-peaks on both sides of the main-peak of its autocorrelation function, causing an ambiguity problem in the BOC signal synchronization. Moreover, the number of side-peaks increases as the value of k becomes larger. The synchronization process of BOC signals consists of two stages: Acquisition and tracking. The acquisition process aligns the locally generated BOC signal with the received signal within a chip duration, and then, the tracking process performs fine synchronization and maintains the synchronized lock point. To solve the ambiguity problem, several unambiguous acquisition schemes [7]-[11] and tracking schemes [12]-[14] have been proposed. In [7]-[9], sideband filtering was used to deal with the ambiguity problem in the BOC signal acquisition; however, these schemes destroy the sharpness of the main-peak of the BOC autocorrelation function, degrading the BOC signal tracking performance severely. In [10], an interesting unambiguous acquisition scheme that maintains the sharp main-peak of the BOC autocorrelation function was proposed combining the correlation between the BOC and PRN signals with the BOC autocorrelation; however, this scheme is applicable to only $\mathrm{BOC}_{\sin}(kn,n)$ signals. In [11], an extended unambiguous acquisition scheme including the scheme in [10] as a special case was proposed. This scheme is applicable to generic $\mathrm{BOC}_{\sin}(kn,n)$ signals; however, its extension to generic $\mathrm{BOC}_{\cos}(kn,n)$ signals is not straightforward since it uses the designed local signals for $\mathrm{BOC}_{\sin}(kn,n)$ signals. In [12]-[14], on the other hand, unambiguous tracking schemes have been proposed by employing side-peak cancellation techniques in a delay lock loop, focusing on the tracking process of the BOC signals, and they are applicable to $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$ signals. In this paper, we focus on acquisition of the BOC signals.

In this paper, a novel unambiguous acquisition scheme applicable to both $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$ signals is proposed based on a recombination of the sub-correlations making up the BOC autocorrelation, which is found to remove the side-peaks of the BOC autocorrelation completely, while keeping the sharp shape of the main-peak, and also, to offer a performance improvement over the scheme in [11] in terms of the correlation function, receiver operating characteristic (ROC) curves (which is the probability of detection P_D as a function of the probability of false alarm P_{FA}), and mean acquisition time (MAT) (which is the time that elapses prior to acquisition on the average

*Corresponding author

 ISBN: 978-1-61208-305-6



(a) When $k = 1$ (b) When $k = 2$

Figure 1. BOC autocorrelation and the associated sub-correlations when (a) k = 1 and (b) k = 2 in the absence of noise.

and the ultimate performance metric of interest for signal acquisition).

II. ANALYSIS OF BOC AUTOCORRELATION FUNCTION

The BOC signal $b(t)$ can be expressed as:

$$b(t) = \sqrt{P}\sum_{i=-\infty}^{\infty} c_i\, p_{T_c}(t - iT_c)\, d_{\lfloor iT_c/T \rfloor}(t)\, s(t), \qquad (1)$$

where P is the signal power, $c_i \in \{-1,1\}$ is the i th chip of a PRN code with a period of T, T_c is the PRN code chip period, $p_{T_c}(t)$ is the PRN code waveform defined as a unit rectangular pulse over $[0, T_c)$, $d_{\lfloor iT_c/T \rfloor}(t)$ is the navigation data, where $d_x(t)$ is the x th navigation data and $\lfloor x \rfloor$ is the largest integer not larger than x, and

$$s(t) = \begin{cases} \sum_{u=0}^{2k-1} (-1)^u\, p_{T_s}(t - iT_c - uT_s), \text{for BOC}_{\sin}(kn,n) \\ \sum_{u=0}^{4k-1} (-1)^{\lceil \frac{u}{2} \rceil}\, p_{\frac{T_s}{2}}(t - iT_c - \frac{uT_s}{2}), \text{for BOC}_{\cos}(kn,n) \end{cases} \qquad (2)$$

is the square wave sub-carrier, where T_s is the sub-carrier pulse duration of $T_c / 2k = 1/(2kn \times 1.023\,\text{MHz})$, $p_{T_s}(t)$ is the unit rectangular sub-carrier pulse waveform over $[0, T_s)$, and $\lceil x \rceil$ is the smallest integer not less than x. In this paper, focusing on the problem of ambiguity due to side-peaks, we assume that there is a pilot channel for acquisition [15] so that no data modulation is present during acquisition (i.e., $d_{\lfloor iT_c/T \rfloor}(t) = 1$ for all i), and do not consider the effect of the secondary code. Then, considering that the PRN code period T is generally much larger than the PRN code chip period T_c and the out-of-phase autocorrelation of a PRN code is designed to be as low as possible for easy signal acquisition, we can obtain the correlation (normalized to the signal power) between the received and locally generated BOC signals as [16]:

$$\begin{aligned} R_{\sin}^{k}(\tau) &= \frac{1}{PT}\int_0^T (b(t-\tau) + w(t))\, b(t)\, dt \\ &\cong \sum_{u=0}^{2k-1} \Big(\frac{1}{T_c}\sum_{v=0}^{2k-1} (-1)^{u+v}\, \Lambda_{T_s}(\tau - (u-v)T_s) + w_{\sin}^{u}\Big) \end{aligned} \qquad (3)$$

for $\text{BOC}_{\sin}(kn, n)$ and

$$R_{\cos}^{k}(\tau) \cong \sum_{u=0}^{4k-1} \Big(\frac{1}{T_c}\sum_{v=0}^{4k-1} (-1)^{\lceil \frac{u}{2} \rceil + \lceil \frac{v}{2} \rceil}\, \Lambda_{\frac{T_s}{2}}(\tau - (u-v)\frac{T_s}{2}) + w_{\cos}^{u}\Big) \qquad (4)$$

for $\text{BOC}_{\cos}(kn, n)$, where τ is the phase difference between the received and locally generated BOC signals, $w(t)$ is the additive white Gaussian noise (AWGN) process with mean zero and one-sided noise power spectral density N_0,

$w_{\sin}^{u} = \frac{1}{\sqrt{PT}}\int_0^T \sum_{i=-\infty}^{\infty} (-1)^u c_i\, p_{T_s}(t - iT_c - uT_s)\, w(t)\, dt$,

$w_{\cos}^{u} = \frac{1}{\sqrt{PT}}\int_0^T \sum_{i=-\infty}^{\infty} (-1)^{\lceil \frac{u}{2} \rceil} c_i\, p_{\frac{T_s}{2}}(t - iT_c - \frac{uT_s}{2})\, w(t)\, dt$, and

$$\Lambda_x(\tau) = \begin{cases} x - |\tau|, & |\tau| \le x, \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

is the triangular function of height x and area x^2. Denoting the terms $(\frac{1}{T_c}\sum_{v=0}^{2k-1} (-1)^{u+v}\, \Lambda_{T_s}(\tau - (u-v)T_s) + w_{\sin}^{u})$ and $(\frac{1}{T_c}\sum_{v=0}^{4k-1} (-1)^{\lceil \frac{u}{2} \rceil + \lceil \frac{v}{2} \rceil}\, \Lambda_{\frac{T_s}{2}}(\tau - (u-v)\frac{T_s}{2}) + w_{\cos}^{u})$ in (3) and (4) by $R_{\sin}^{k,u}(\tau)$ and $R_{\cos}^{k,u}(\tau)$, respectively, we can re-write $R_{\sin}^{k,u}(\tau)$ and $R_{\cos}^{k,u}(\tau)$ as:

$$\begin{aligned} R_{\sin}^{k,u}(\tau) &= \frac{1}{T_c}\sum_{v=0}^{2k-1} (-1)^{u+v}\, \Lambda_{T_s}(\tau - (u-v)T_s) + w_{\sin}^{u} \\ &= \sum_{l=0}^{L-1} \frac{1}{PT}\int_{(2kl+u-1)T_s}^{(2kl+u)T_s} r(t)\, b(t)\, dt, \end{aligned} \qquad (6)$$

and similarly,

$$R_{\cos}^{k,u}(\tau) = \sum_{l=0}^{L-1} \frac{1}{PT} \int_{(\frac{4kl+u}{2}-\frac{1}{2})T_s}^{(\frac{4kl+u}{2})T_s} r(t)b(t)dt, \qquad (7)$$

where $r(t) = b(t-\tau) + w(t)$ and L is a correlation length and would be generally limited to be equal to or less than the PRN code period (normalized to T_c) due to some constraints such as the frequency error, data modulation, and secondary code. From (6) and (7), we can see that $R_{\sin}^{k,u}(\tau)$ and $R_{\cos}^{k,u}(\tau)$ are sub-correlations making up the correlations (3) and (4), respectively, and which are shown for $k = 1$ and $k = 2$ in the absence of noise in Fig. 1. From the figure, we can see that the main-peaks of the sub-correlations are coherently combined through the summation of the sub-correlations, thus forming the sharp main-peak of the BOC autocorrelation, and on the other hand, the sub-peaks of the sub-correlations are irregularly spread around the main-peaks, and thus, the summation of the sub-correlations results in the multiple side-peaks of the BOC autocorrelation. Another important observation is that the number of the side-peaks increases as *k* increases. From this observation, it is expected that the acquisition performance is degraded as *k* increases. In the next section, we propose a novel unambiguous acquisition scheme, removing the side-peaks completely through a recombination of the sub-correlations.

III. PROPOSED UNAMBIGUOUS ACQUISITION SCHEME

From Fig. 1, we can clearly observe that $R_{\sin}^{k,0}(\tau)$ and $R_{\sin}^{k,2k-1}(\tau)$ and $R_{\cos}^{k,0}(\tau)$ and $R_{\cos}^{k,4k-1}(\tau)$ are symmetric with respect to $\tau = 0$ and have only a single overlapped peak at $\tau = 0$ for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$, respectively. Thus, if the two sub-correlations are summed, a main-peak with a larger magnitude (than that of the main-peak of a sub-correlation) is obtained without increasing the magnitudes of the side-peaks, and on the other hand, the difference between the two sub-correlations yields side-peaks only, whose magnitudes and positions are the same as those of the side-peaks in the sum of the two sub-correlations. Thus, the difference between the two sub-correlations might be used to remove the side-peaks in the sum of the two sub-correlations, leaving only the main-peak. This observation is the key motivation of the proposed scheme.

Since the side-peaks in the sum and difference of the two sub-correlations are out-of-phase and in-phase at $\tau < 0$ and $\tau > 0$, respectively, however, we cannot remove the side-peaks in the sum of the two sub-correlations completely through the subtraction between the sum and difference of the two sub-correlations. To align the phases of the side-peaks in the sum and difference of the two sub-correlations, thus, we use the sum of the absolute values of the two sub-correlations, obtaining the side-peaks with the same slopes as those of the side-peaks in the absolute difference of the two sub-correlations. Fig. 2 shows that the subtraction of the absolute difference of the two sub-correlations from the sum of the absolute values of the two sub-correlations yields an

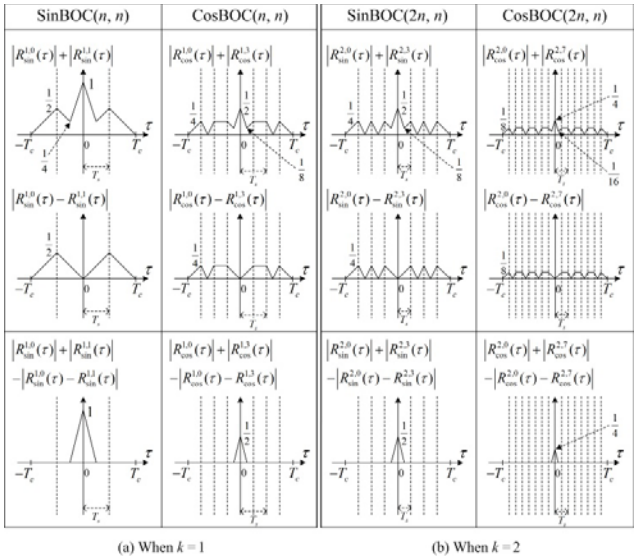


Figure 2. Unambiguous correlation functions obtained through subtraction between the sub of absolute values and absolute difference of $R_{\sin}^{k,0}(\tau)$ $R_{\sin}^{k,2k-1}(\tau)$ and $R_{\cos}^{k,0}(\tau)$ and $R_{\cos}^{k,4k-1}(\tau)$ when (a) *k* = 1 and (b) *k* = 2 in the absence of noise.

unambiguous correlation function with a single main-peak and no side-peak.

Since the unambiguous correlation function is generated with several absolute operations, it may suffer from undesired noise enhancement. To alleviate the noise enhancement while maintaining the signal part (i.e., the main-peak), we first multiply the unambiguous correlation function with each of the sub-correlations, and then, sum the product results together. Since the noise random variables $\{w_{\sin}^{u}\}_{u=0}^{2k-1}$ ($\{w_{\cos}^{u}\}_{u=0}^{4k-1}$) in $\{R_{\sin}^{k,u}(\tau)\}_{u=0}^{2k-1}$ ($\{R_{\cos}^{k,u}(\tau)\}_{u=0}^{4k-1}$) are independent from each other under the AWGN-limited satellite environment, the product sum of $\{R_{\sin}^{k,u}(\tau)\}_{u=0}^{2k-1}$ ($\{R_{\cos}^{k,u}(\tau)\}_{u=0}^{4k-1}$) will average out the noise, and moreover, the main-peak magnitude of the unambiguous correlation function can be maintained after the product sum, since all sub-correlations have an equal magnitude of 1 / (2*k*) and 1 / (4*k*), the inverse of the number of the sub-correlations, for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$, respectively.

From the above discussions, the proposed unambiguous correlation function can be expressed as:

$$R_{\sin}^{k,\mathrm{proposed}}(\tau) = \sum_{u=0}^{2k-1} R_{\sin}^{k,u}(\tau)(|R_{\sin}^{k,0}(\tau)| + |R_{\sin}^{k,2k-1}(\tau)| - |R_{\sin}^{k,0}(\tau) - R_{\sin}^{k,2k-1}(\tau)|) \qquad (8)$$

for $\mathrm{BOC}_{\sin}(kn,n)$ and

$$R_{\cos}^{k,\mathrm{proposed}}(\tau) = \sum_{u=0}^{4k-1} R_{\cos}^{k,u}(\tau)(|R_{\cos}^{k,0}(\tau)| + |R_{\cos}^{k,4k-1}(\tau)| - |R_{\cos}^{k,0}(\tau) - R_{\cos}^{k,4k-1}(\tau)|) \qquad (9)$$

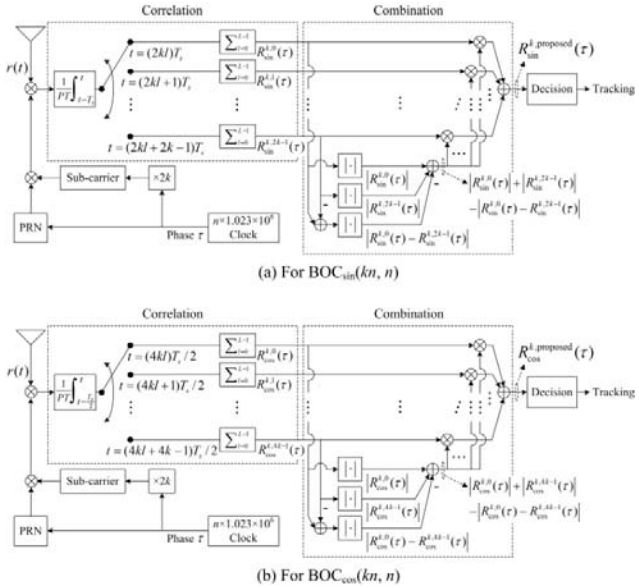


Figure 3. The baseband receiver structure of the proposed unambiguous acquisition scheme for (a) $\mathrm{BOC}_{\sin}(kn,n)$ and (b) $\mathrm{BOC}_{\cos}(kn,n)$.

for $\mathrm{BOC}_{\cos}(kn,n)$.

Fig. 3 shows the baseband receiver structure of the proposed unambiguous acquisition scheme for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$. The received BOC signal $r(t)$ is first multiplied with the locally generated PRN code and sub-carrier, and then, integrated and sampled every T_s and $T_s/2$ seconds for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$, respectively. Subsequently, sub-correlation values are obtained by summing L samples per sub-correlation, and then, combined according to (8) and (9) to produce a decision variable based on $R_{\sin}^{k,\mathrm{proposed}}(\tau)$ and $R_{\cos}^{k,\mathrm{proposed}}(\tau)$, respectively. Finally, decision variables corresponding to possible phases in the uncertainty region are collected and the process is transferred to the tracking stage with the phase corresponding to the largest of the decision variables. It should be noted that the proposed scheme needs only a single correlator since each sub-correlation is sequentially obtained by sampling the single correlator output every $T_s\ (=T_c/2k)$ seconds and every $T_s/2\ (=T_c/4k)$ seconds for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$ signals, respectively, as shown in Fig. 3.

IV. NUMERICAL RESULTS

In this section, the proposed unambiguous acquisition scheme is compared with the unambiguous acquisition scheme in [11] called the general removing ambiguity via side-peak suppression (GRASS) in terms of the correlation function, ROC curves, MAT. In comparisons, we assume the following parameters: Galileo E1-C PRN code of $T = 4092$chips [4], correlation length of $L = 1023$, and a search step size of T_s and $T_s/2$ for the sine-phased and cosine-phased BOC signals, respectively.

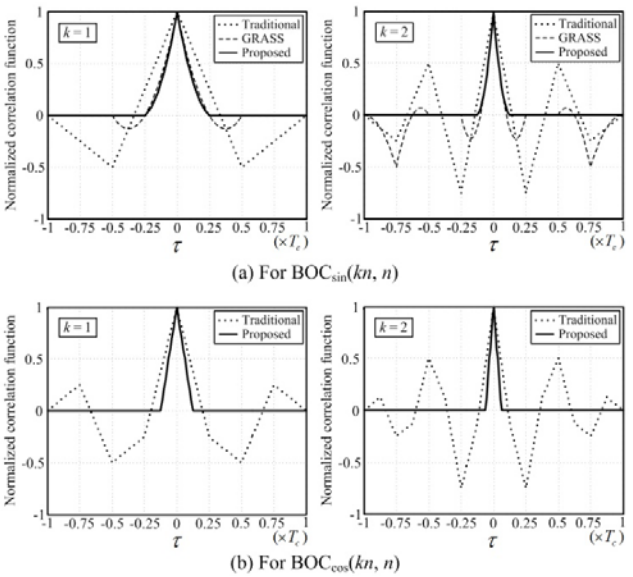


Figure 4. Normalized correlation functions of the proposed, GRASS, traditional BOC schemes for (a) $\mathrm{BOC}_{\sin}(kn,n)$ and (b) $\mathrm{BOC}_{\cos}(kn,n)$ when k = 1 and k = 2 in the absence of noise.

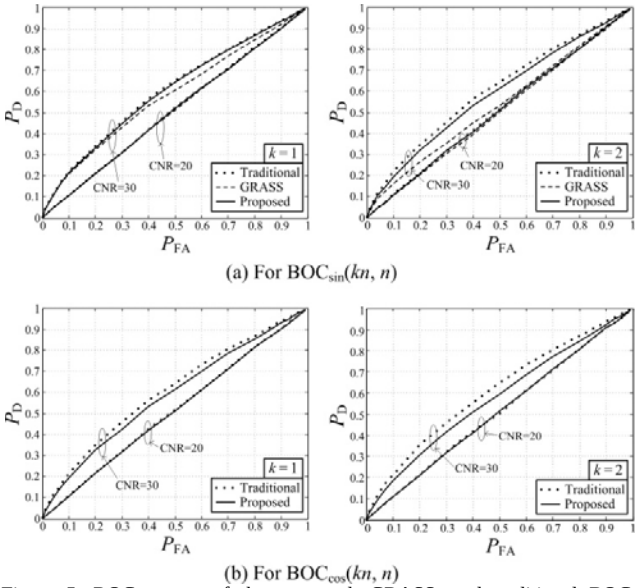


Figure 5. ROC curves of the proposed, GRASS, and traditional BOC schemes for (a) $\mathrm{BOC}_{\sin}(kn,n)$ and (b) $\mathrm{BOC}_{\cos}(kn,n)$ when k = 1 and k = 2.

Fig. 4 shows the normalized correlation functions of the proposed, GRASS, and traditional BOC schemes for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$ when $k=1$ and k = 2 in the absence of noise, where the GRASS correlation function is not shown for $\mathrm{BOC}_{\cos}(kn,n)$ since it is dedicated to the sine-phased BOC signals only, and the traditional BOC autocorrelation is also shown as a reference. From the figure, unlike the GRASS and traditional BOC schemes, the proposed scheme is clearly observed to remove the side-peaks completely for both $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$ regardless of the value of k.

 ISBN: 978-1-61208-305-6

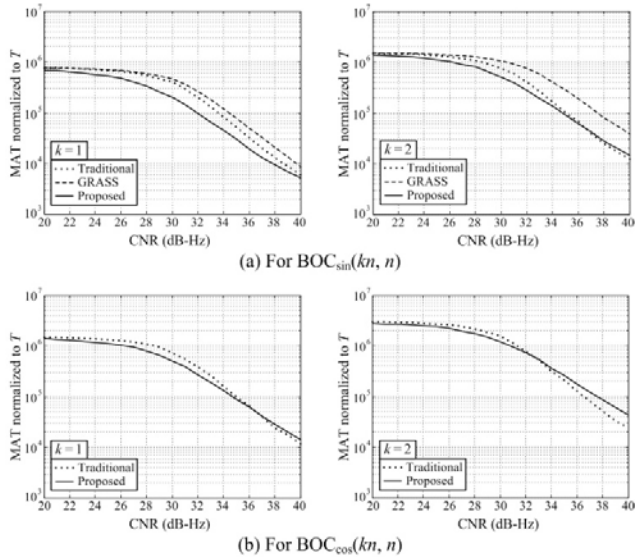


Figure 6. MAT of the proposed, GRASS, and traditional BOC schemes for (a) $\mathrm{BOC}_{\sin}(kn,n)$ and (b) $\mathrm{BOC}_{\cos}(kn,n)$ when $k = 1$ and $k = 2$.

Fig. 5 shows the ROC curves of the proposed, GRASS, and traditional BOC schemes for $\mathrm{BOC}_{\sin}(kn,n)$ and $\mathrm{BOC}_{\cos}(kn,n)$ when $k=1$ and $k = 2$. The three schemes are compared with the results in the AWGN environments when the carrier-to-noise ratios (CNRs) are 20 and 30 dB-Hz, where the CNR is defined as P/N_0 (dB-Hz). From the figure, it is observed that the relative performance of the proposed and traditional BOC schemes is the same, whereas the proposed scheme offers a performance improvement over the GRASS scheme.

As shown in Fig. 5, it is seen that the traditional BOC scheme provides the best ROC performances; however, it should be noted the probabilities of detection and false alarm of the ROC curve are associated with only the main-peak of the correlation function, i.e., the ROC curve does not reflect the effect of the side-peaks elimination. Thus, we compare the MAT performances of the three schemes as shown in Fig. 6, where the penalty time and the probability of false alarm are set to $4T$ and 10^{-3}, respectively. As shown in the figures, the proposed scheme performs better than the GRASS and traditional BOC schemes in the CNR range 20 ~ 40 dB-Hz of practical interest.

V. CONCLUSION

In this paper, we have proposed a novel unambiguous BOC acquisition scheme for GNSSs. First, we have analyzed that the BOC autocorrelation is made up of the sum of several sub-correlations, and then, observed that the irregular shapes of the sub-correlations cause the side-peaks in the BOC autocorrelation. Then, we have proposed an unambiguous correlation function with no side-peak via a recombination of the sub-correlations. From numerical results, it has been observed that the proposed scheme removes the side-peaks completely providing a performance improvement over the GRASS and traditional BOC schemes in terms of the ROC curves and MAT.

ACKNOWLEDGMENT

This research was supported by the National Research Foundation (NRF) of Korea under Grant 2012R1A2A2A01045887 with funding from the Ministry of Science, ICT & Future Planning (MSIP), Korea, by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency under Grant NIPA-2013-H0301-13-1005 with funding from the MSIP, Korea, and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

REFERENCES

[1] E. Kaplan and C. Hegarty, Understanding GPS: Principles and Applications, 2nd ED., Norwood: Artech House, 2006.

[2] E. S. Lohan, A. Lakhzouri, and M. Renfors, "Binary-offset-carrier modulation techniques with applications in satellite navigation systems," Wireless Commun. Mobile Computing, vol. 7, Aug. 2007, pp. 767-779.

[3] M. Zahidul, H. Bhuiyan, E. S. Lohan, and M. Renfors, "Code tracking algorithms for mitigating multipath effects in fading channels for satellite-based positioning," Eurasip Journal on Advances in Signal Process., vol. 2008, Jan. 2008, pp. 1-17.

[4] J. A. Avila-Rodriguez, "On generalized signal waveforms for satellite navigation," Ph.D. dissertation, Dept. Aer. Engineer., University of Munich, Munich, Germany, 2008.

[5] W. Liu, G. Du, X. Zhan, and C. Zhai, "MSK-binary coded symbol modulations for global navigation satellite systems," IEICE Electron. Express, vol. 7, Mar. 2010, pp. 421-427.

[6] J. Wu and A. G. Dempster, "Applying a BOC-PRN discriminator to cosine phased BOC(f_s, f_c) modulation," Electron. Lett., vol. 45, June 2009, pp. 689-690.

[7] N. Martin, V. Leblond, G. Guillotel, and V. Heiries, "BOC(x,y) signal acquisition techniques and performances," Proc. ION GPS/GNSS, Omnipress , Sep. 2003, pp. 188-198.

[8] A. Burian, E. S. Lohan, V. Lehtinen, and M. K. Renfors, "Complexity considerations for unambiguous acquisition of Galileo signals," Proc. Workshop on Positioning, Navig., and Commun., Shaker Publishing, Mar. 2006, pp. 65-74.

[9] E. S. Lohan, A. Burian, and M. Renfors, "Low-complexity unambiguous acquisition methods for BOC-modulated CDMA signals," Int. Journal of Satell. Commun. Networking, vol. 26, Nov.-Dec. 2008, pp. 503-522.

[10] O. Julien, C. Macabiau, M. E. Cannon, and G. Lachapelle, "ASPeCT: unambiguous sine-BOC(n,n) acquisition/tracking technique for navigation applications," IEEE Trans. Aer. Electron. Syst., vol. 43, Jan. 2007, pp. 150-162.

[11] Z. Yao, M. Lu, and Z. Feng, "Unambiguous sine-phased binary offset carrier modulated signal acquisition technique," IEEE Trans. wireless Commun., vol. 9, Feb. 2010, pp. 577-580.

[12] S. Kim, S. Yoo, S. Yoon, and S. Y. Kim, "A novel unambiguous multipath mitigation scheme for BOC(kn, n) tracking in GNSS," Proc. 7th IEEE Int. Symp. App. Internet (SAINT), IEEE CS Press, Jan. 2007, pp. 57-60.

[13] Y. Lee, Y. Lee, T. Yoon, C. Song, S. Kim, and S. Yoon, "AltBOC and CBOC correlation functions for GNSS signal synchronization," Springer-Verlag Lecture Notes in Computer Science, vol. 5593, June 2009, pp. 325-334.

[14] Y. Lee, D. Chong, I. Song, S. Y. Kim, G.-I. Jee, and S. Yoon, "Cancellation of correlation side-peaks for unambiguous BOC

 ISBN: 978-1-61208-305-6

signal tracking," IEEE Commun. Lett., vol. 16, May 2012, pp. 569-572.

[15] F. D. Nunes, M, G. Sousa, and J. M. N. Leitao, "Gating functions of for multipath mitigation in GNSS BOC signals," IEEE Trans. Aer. Electron. Syst., vol. 43, July 2007, pp. 951-964.

[16] E. S. Lohan, A. Lakhzouri, and M. Renfors, "Feedforward delay estimators in adverse multipath propagation for Galileo and modernized GPS signals," Eurasip Journal on Applied Signal Process., vol. 2006, Jan. 2006, pp. 1-19.

 ISBN: 978-1-61208-305-6

A Novel Cognitive Engine Based on Genetic Algorithm

Keunhong Chae, Youngseok Lee, and Seokho Yoon*
College of Information and Communication Engineering
Sungkyunkwan University
Suwon, Korea
e-mail: {chae0820, fortrtwo, and *syoon}@skku.edu

***Abstract*—In this paper, we propose a novel cognitive engine based on genetic algorithm (GA). Unlike conventional GA-based cognitive engines, the proposed cognitive engine takes the frequency band of the secondary user as one of the transmission parameters to be optimized, allowing the proposed cognitive engine to choose the optimal frequency band among the vacant bands detected via the spectrum sensing. Numerical results demonstrate that the proposed cognitive engine well optimizes the transmission parameters for a given transmission scenario.**

Keywords-cognitive engine; genetic algorithm; transmission parameter; optimization

I. INTRODUCTION

The increasing demand for high-speed multimedia services has led to the advent of various wideband wireless communication systems including long term evolution (LTE), IEEE 802.16, digital video broadcasting (DVB), and Wi-Fi. As the number of the wideband wireless communications and associated subscribers increases, the spectrum deficiency problem is inevitable since the frequency spectrum is a limited resource. To resolve the problem, the dynamic spectrum access (DSA) technique, which opportunistically utilizes an underutilized frequency band, has been proposed by virtue of the software defined radio (SDR) capable of tuning its transmission parameters [1].

A secondary user (SU) observes the surrounding environments, and subsequently, adjusts its transmission parameters (e.g., the transmit power, modulation index, and transmission bandwidth) based on the observation. Specifically, the SU first determines if a primary user (PU) is utilizing the spectrum band of interest via a spectrum sensing [2]. Then, the transmission parameters of the SU are optimized by an intelligent signal processing unit referred to as a cognitive engine. The implementation of the cognitive engine has been studied mainly based on the artificial intelligence (AI) techniques such as the genetic algorithm (GA), expert systems, neural networks, and case-based reasoning [3]. Especially, the GA-based cognitive engine has attracted much attention since it is capable of self-evolution as the human cognition process unlike other AI-based cognitive engines [4].

In wideband wireless environments, a wideband spectrum is generally interpreted as a set of multiple narrowbands. Thus, after the vacant narrowbands are identified via the spectrum sensing, the cognitive engine should choose an optimal narrowband out of the vacant ones. Thus, in this paper, we consider the frequency band of an SU as one of the transmission parameters to be optimized, and subsequently, propose a cognitive engine by designing a multiple objective fitness function that includes the frequency band of the SU as a transmission parameter, and then, applying the fitness function to the genetic algorithm.

The rest of this article is organized as follows. Section II introduces the transmission parameters and the cognitive engine system model. Then, in Section III, a multiple objective fitness function is proposed taking the frequency band of SU as a transmission parameter. Section IV demonstrates that a cognitive engine employing the proposed fitness function appropriately optimizes the transmission parameters, and finally, Section V concludes the paper.

II. RELATED WORK

Several studies on the GA-based cognitive engine have been researched focusing on how to optimize the transmission parameters of the SU [5]-[10]. In [5] and [6], an initial version of a GA-based cognitive engine was implemented as a hardware test-bed proving its usefulness as a cognitive engine. To deal with various transmission scenarios, in [7], a multiple objective fitness function is designed as a weighted sum of single objective fitness functions. Recently, to optimize the multiple objective fitness function, cognitive engines have been proposed by employing various evolutionary algorithms such as artificial bee colony algorithm, ant colony optimization, and Biogeography-based optimization instead of GA [8]-[10]. However, the conventional researches were focused on the transmission parameter optimization after the spectrum assignments have been determined, and thus, the frequency band of an SU has not been optimized as a transmission parameter.

III. SYSTEM MODEL

Transmission parameters are the variables to be optimized based on information in environment parameters (e.g., the noise density and the value of the test statistic used in the spectrum sensing). In this paper, we consider the following transmission parameters: the transmit power P_s of SU, modulation index M, bandwidth B_s of the SU signal, and index k of the frequency band that is detected as a vacant band via the spectrum sensing.

To optimize the transmission parameters using GA, we first design a structure of the chromosome as a bit stream

*Corresponding author

 ISBN: 978-1-61208-305-6

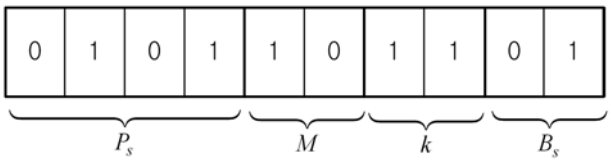


Figure 1. The structure of a chromosome representing the transmission parameters.

representing the values of the transmission parameters. For example, we can design a chromosome as a bit stream with a length of 10, where 4, 2, 2, and 2 bits are used to represent the values of P_s, M, k, and B_s, respectively, as shown in Fig. 1. In this case, we are dealing with 16, 4, 4, and 4 candidates for P_s, M, k, and B_s, respectively. Then, a multiple objective fitness function is defined to measure the desirability of a solution for a transmission scenario of interest. The multiple objective fitness function f can be expressed as

$$f = a_1 f_1 + a_2 f_2 + ... + a_L f_L, \quad (1)$$

where $\{f_l\}_{l=1}^{L}$ are the single objective fitness functions and $\{a_l\}_{l=1}^{L}$ denote the weight values for $\{f_l\}_{l=1}^{L}$ and $\sum_{l=1}^{L} a_l = 1$ [7]. A transmission scenario determines the value of weights $\{a_l\}_{l=1}^{L}$ by assigning a higher (lower) weight to the single objective fitness function with higher (lower) priority. Finally, the GA provides an optimum solution maximizing the designed fitness function. Specifically, a fitness value of an initial set of the transmission parameter values is calculated, and then, searches for an optimum set by using the selection, crossover, and mutation operations.

IV. PROPOSED FITNESS FUNCTION

The procedure to design a multiple objective fitness function in a form of (1) is to determine single objective fitness functions and associated weight values. We first design single objective fitness functions that affect the data transmission performance of the SU including bit error rate (BER) and throughput. The smaller value of BER guarantees the more reliable performance of the SU, and thus, a single objective fitness function f_{BER} is designed as

$$f_{\text{BER}} = \frac{\log_{10}(0.5) - \log_{10}(P_b)}{\log_{10}(0.5) - \log_{10}(P_{b,\min})}, \quad (2)$$

where P_b is the BER and $P_{b,\min}$ is the minimum value of $\{P_b\}$ for the given candidates of the transmission parameters. When the BER P_b is expressed in terms of E_b/N_0, where E_b and N_0 denote the bit energy and the noise density, respectively, P_b can also be expressed in terms of the transmission parameters by substituting $\frac{2P_s}{B_s \times \log_2(M) \times N_0}$ for E_b/N_0. On the other hand, the throughput of the data transmission is proportional to the modulation index M, thus, a single objective fitness function $f_{\text{throughput}}$ can be expressed as

$$f_{\text{throughput}} = \frac{\log_2(M) - \log_2(M_{\min})}{\log_2(M_{\max}) - \log_2(M_{\min})}, \quad (3)$$

where $M_{\max}$ and $M_{\min}$ are the maximum and minimum values of $\{M\}$, respectively. The function is maximized (minimized) when $M = M_{\max}$ ($M = M_{\min}$).

It is also desired to reduce interference to the PU signal, which depends on the power P_s and bandwidth B_s of the SU signal. Thus, we design a single fitness function $f_{\text{interference}}$ as

$$f_{\text{interference}} = 1 - \frac{1}{2}\left\{\left(\frac{P_s - P_{s,\min}}{P_{s,\max} - P_{s,\min}}\right) + \left(\frac{B_s - B_{s,\min}}{W(k) - B_{s,\min}}\right)\right\}, \quad (4)$$

where $P_{s,\max}$ and $P_{s,\min}$ are the maximum and minimum values of $\{P_s\}$, respectively, $W(k)$ is the bandwidth of the kth narrowband assigned to the PU, and $B_{s,\min}$ is the minimum value of the bandwidth candidates B_s.

Now, we will discuss how to obtain the optimal value of k and design a single objective fitness function. It is naturally assumed that the test statistic value $T(k)$ and the threshold $\gamma(k)$ for the spectrum sensing of the kth band is known to the cognitive engine, and the bandwidth $W(k)$ is a priori knowledge. Although the candidate narrowbands are detected as a vacant band by the spectrum sensing process, some of the narrowbands may be occupied by the PU signal due to the missed detection of the spectrum sensing. Thus, we design a term $\left(\frac{D(k) - D_{\min}}{D_{\max} - D_{\min}}\right)$, where $D(k) = \gamma(k) - T(k)$, and $D_{\max}$ and $D_{\min}$ are the maximum and minimum values of $\{D(k)\}$, respectively, based on the observation that the frequency band is more likely to be vacant when the difference between $\gamma(k)$ and $T(k)$ is a larger value. It is noteworthy that $D(k) > 0$ since the narrowbands of interest are already detected as a vacant band (i.e., $\gamma(k) > T(k)$) in the spectrum sensing process. Moreover, to choose a wide frequency band and to fully use the selected frequency band, we also design two terms $\left(\frac{W(k) - W_{\min}}{W_{\max} - W_{\min}}\right)$ and $\left(\frac{B_s - B_{s,\min}}{W(k) - B_{s,\min}}\right)$, where $W_{\max}$ and $W_{\min}$ are the maximum and minimum values of $\{W(k)\}$, respectively, and $B_{s,\max}$ is the maximum

 ISBN: 978-1-61208-305-6

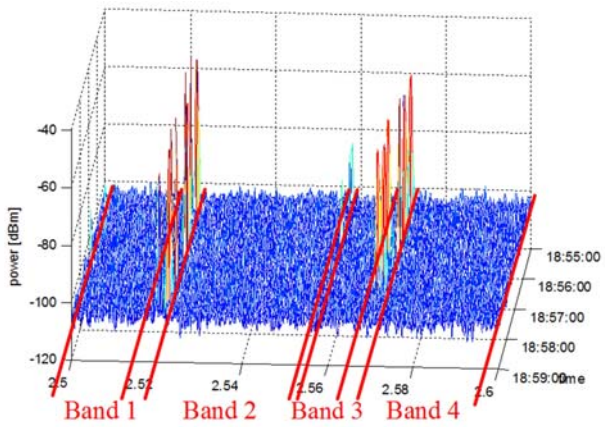


Figure 2. The frequency spectrum of [250 MHz, 260 MHz] bands.

value of the bandwidth candidates B_s of the SU. Normalizing and combining the designed terms, we propose a single objective fitness function f_{band} as

$$f_{\text{band}} = \frac{1}{3}\left(\frac{D(k)-D_{\min}}{D_{\max}-D_{\min}}\right)+\frac{1}{3}\left(\frac{W(k)-W_{\min}}{W_{\max}-W_{\min}}\right) + \frac{1}{3}\left(\frac{B_s-B_{s,\min}}{W(k)-B_{s,\min}}\right). \quad (5)$$

In summary, the function f_{band} is designed (i) to maximize the probability that the chosen band is vacant, (ii) to choose a band with a larger bandwidth, and (iii) to transmit the SU signal with a larger bandwidth.

Finally, we propose a multiple objective fitness function as

$$f_{\text{WB}} = w_1 f_{\text{band}} + w_2 f_{\text{BER}} + w_3 f_{\text{throughput}} + w_4 f_{\text{interference}}, \quad (6)$$

where $\{w_l\}_{l=1}^4$ are the weight values for fitness functions for f_{band}, f_{BER}, $f_{\text{throughput}}$, and $f_{\text{interference}}$, and $\sum_{l=1}^4 w_l = 1$.

V. NUMERICAL RESULTS

In this section, we explain the cognitive engine simulator that we have developed and show the results on the transmission parameter optimization. We measured a frequency spectrum of [250 MHz, 260 MHz] bands at the top of a mountain and used the measured data as the input of the simulator. The spectrum is shown in Fig. 2, where four spectrum bands (Band 1 ~ Band 4) are detected as the vacant narrowbands. The simulator is developed using Matlab graphic user interface (GUI) programming and its main screen is shown in Fig. 3.

For simulations, we assume the following parameters: a chromosome with a length of 10 bits, where 4, 2, 2, and 2 bits are used to represent the values of P_s, M, k, and B_s,

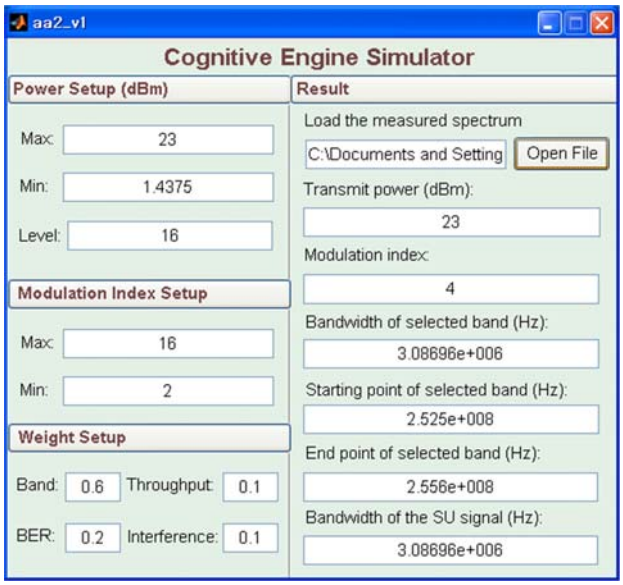


Figure 3. The cognitive engine GUI simulator.

respectively. The candidates for P_s, M, k, and B_s are set to represent $\{\frac{23}{16}, \frac{2\times 23}{16}, \ldots, 23\}$ dBm, $\{2,4,8,16\}$, {Band 1, Band 2, Band 3, Band 4}, and $\{10,100,500,W(k)/10^3\}$ kHz, respectively. The noise density N_0 is calculated as the power spectral density of a frequency band with the lowest power over the spectrum range of [200 MHz, 300 MHz], then, the threshold for the spectrum sensing is determined to satisfy the false alarm probability of 0.01. The spectrum sensing is performed via the energy detector.

For the transmission scenarios, we first consider the simplest case that $\bar{w} = [w_1, w_2, w_3, w_4] = [1,0,0,0]$ to verify the simulator, and subsequently, we demonstrate the results for a scenario with the weight vector $[0.6, 0.1, 0.2, 0.1]$ as an example. Fig. 4 shows (a) the fitness value and (b) the result solution of the simulator when $\bar{w} = [1,0,0,0]$. From the figure, we can see that the fitness value becomes saturated as the generation increases. Also, the 7th and 8th bits of the chromosome are '01' and the 9th and 10th bits are '11' representing that Band 2 (the largest vacant band) is chosen as the frequency band and the SU uses the whole spectrum of Band 2. However, the transmit power (the 1st-4th bits) and the modulation index (the 5th and 6th bits) are randomly selected by the GA since the fitness function f_{band} is not a function of P_s and M.

Fig. 5 shows (a) the fitness value and (b) the result solution of the simulator when $\bar{w} = [0.6, 0.2, 0.1, 0.1]$. Since the weight value for f_{band} is the largest, Band 2 is chosen as the frequency band and the SU uses the whole spectrum of Band 2 as in the case that $\bar{w} = [1,0,0,0]$; however, for the transmit power P_s and the modulation index M, the maximum power of 23 dBm and QPSK modulation is

 ISBN: 978-1-61208-305-6

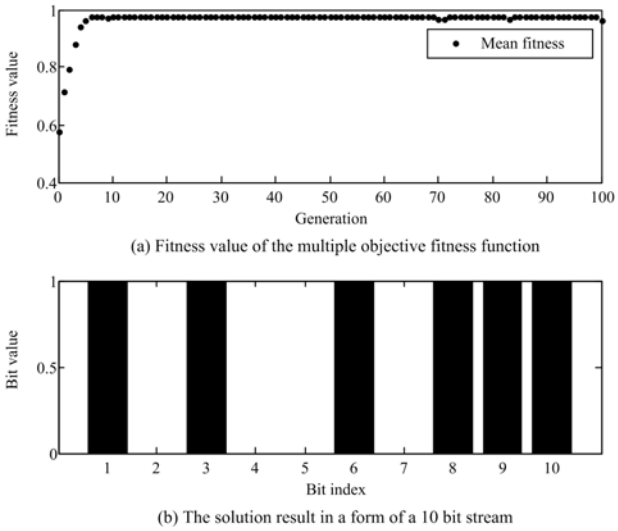


Figure 4. Simulation results when $\bar{w} = [1, 0, 0, 0]$.

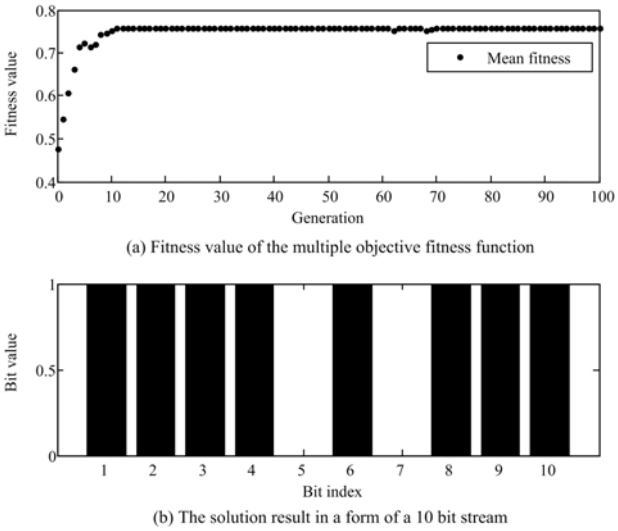


Figure 5. Simulation results when $\bar{w} = [0.6, 0.1, 0.2, 0.1]$.

selected considering the fact that the weight value for f_{BER} is the second largest.

VI. Conclusion and Future Work

In this paper, we have proposed a GA-based cognitive engine suitable for the wideband wireless communications. Including the frequency band of the SU as a transmission parameter to be optimized, we have designed a multiple objective fitness function that measures the desirability of a solution, and then, applied the fitness function to the GA. From numerical results, it has been confirmed that the proposed cognitive engine appropriately optimizes the transmission parameters for a given weight vector describing the transmission scenario. To implement a cognitive engine, it is also required to optimize the weight vector $\bar{w}$ as well as the transmission parameters, which is our future research topic.

Acknowledgment

This research was supported by the National Research Foundation (NRF) of Korea under Grant 2012R1A2A2A01045887 with funding from the Ministry of Science, ICT & Future Planning (MSIP), Korea, by the Information Technology Research Center (ITRC) program of the National IT Industry Promotion Agency under Grant NIPA-2013-H0301-13-1005 with funding from the MSIP, Korea, and by National GNSS Research Center program of Defense Acquisition Program Administration and Agency for Defense Development.

References

[1] P. Yadav, S. Chatterjee, and P. P. Bhattacharya, “A survey on dynamic spectrum access techniques in cognitive radio,” Int. J. Next-Generation Networks, vol. 4, Dec. 2012, pp. 27-46.

[2] T. Yüech and H. Arslan, “A survey of specturm sensing algorithms for cognitive radio applications,” IEEE Commun. Surveys & Tutorials, vol. 11, Mar. 2009, pp. 116-127.

[3] D. Xu, L. Ying, and W. S. Qun, “Design and implementation of a cognitive engine functional architecture,” Chinese Science Bulletin, vol. 57, Oct. 2009, pp. 3698-3704.

[4] C. J. Rieser, “Biologically inspired cognitive radio engine model utilizing distributed genetic algorithms for secure and robust wireless communications and networking,” Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 2004.

[5] D. Maldonado, B. Le, A. Hugine, T. W. Rondeau, and C. W. Bostian, “Cognitive radio applications to dynamic spectrum allocation,” Proc. IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks, IEEE Press, Nov. 2005, pp. 597-600.

[6] T. Rondeau, B. Le, C. Rieser, and C. W. Bostian, “Cognitive radio with genetic algorithms: Intelligent control of software defined radios,” Proc. Software Defined Radio Forum Tech. Conf., SDR Forum, Nov. 2004, pp. C3-C8.

[7] T. R. Newman, B. A. Barker, A. M. Wyglinski, A. Agah, and J. B. Evans, “Cognitive engine implementation for wireless multicarrier transceivers,” Wirel. Commun. Mob. Comput., vol. 7, May 2007, pp. 1129-1142.

[8] P. M. Pradhan, “Design of cognitive radio engine using artificial bee colony algorithm,” Proc. Int. Conf. Energy, Automation, and Signal, IEEE Press, Dec. 2011, pp. 1-4.

[9] N. Zhao, S. Li, and Z. Wu, “Cognitive radio engine design based on ant colony optimization,” Wirel. Pers. Commun., vol. 65, July 2012, pp. 15-24.

[10] K. Kaur, M. Rattan, and M. S. Patterh, “Biogeography-based optimisation of cognitive radio system,” Int. J. Electron., Mar. 2013, pp. 1-13, DOI:10.1080/00207217.2013.769183.

 ISBN: 978-1-61208-305-6

Multiuser Coded FDM-CPM Systems with MIMO Transmission

Piotr Remlein
Chair of Wireless Communications
Poznan University of Technology
Poznan, Poland
e-mail: remlein@et.put.poznan.pl

Mateusz Jasinski
IT Department
Polish Power Grid– West
Poznan, Poland
e-mail: mateusz.jasinski@pse-operator.pl

Alberto Perotti
CSP-ICT Innovation
Torino, Italy
e-mail: alberto.perotti@csp.it

***Abstract*—Performance of coded Frequency-Division Multiplexed Continuous Phase Modulation (FDM-CPM) systems with Multiple Input Multiple Output (MIMO) transmission is investigated. The system is designed to achieve high spectral efficiency by exploiting the multiplexing gain of MIMO techniques. Moreover, a FDM MultiUser (MU) scheme with tight inter-carrier frequency spacing is used to increase spectral efficiency. It is shown that, using this scheme, significant improvement both in terms of bit error rate and spectral efficiency are obtained when compared to the single-antenna MU scenario. To take advantage of the multiplexing gain of MIMO systems, a Minimum Mean Square Error (MMSE) MIMO detector and a low-complexity iterative algorithm for Inter-Carrier Interference (ICI) cancellation are considered. Numerical simulations have been performed to assess the performance improvement achieved with the proposed frequency-division multiplexed CPM multiuser MIMO system.**

Keywords-Continuous Phase Modulation; Frequency-Division Multiplexed system; inter-carrier interference cancellation; multiuser MIMO receiver

I. INTRODUCTION

In the past few years, several methods have been proposed for MultiUser (MU) detection of CPM signals [1], [2], [3], [4]. In [1] and [2], the use of CPM for MU communication over Additive White Gaussian Noise (AWGN) channels and serially concatenated CPM over Rayleigh fading channels were studied. Multiple Input Multiple Output (MIMO) systems [5] can improve bandwidth efficiency by exploiting the channel spatial diversity thus allowing for a transmission of more data streams simultaneously. MIMO systems are often implemented using Space-Time Codes (STC) to increase the reliability of transmission. STC for MIMO CPM systems were previously proposed with appropriate design criteria in [6], and a soft-decision iterative receiver was described [7]. Hesse et al. [8] introduced a new family of orthogonal STC for CPM. These codes offer good performance and low decoding complexity. A non-binary space-time coded scheme with m-ary CPM was developed in [9].

Nevertheless, optimal MU receivers [1] and the MIMO CPM receivers with STC [6], [7], [9] exhibit significantly higher complexity. MIMO CPM receivers for STC studied in [6-9] do not concern MU scenario. In literature [3], [4] the multiuser Frequency-Division Multiplexed (FDM) systems using CPM over AWGN channel were investigated. In [10], a method of estimating Channel State Information (CSI) has been shown to considerably improve performance of MU FDM-CPM systems. However, MIMO techniques are not considered in [3], [4], [10].

This paper investigates the performance of multiuser coded FDM-CPM systems with MIMO transmission. To ensure system simplicity, the employed MIMO scheme does not use Space-Time Codes. The system achieves a high Spectral Efficiency (SE) and low Bit Error Rate (BER), provided that an appropriate low-complexity iterative algorithm for Inter-Carrier Interference (ICI) cancellation is implemented in the MU receiver.

The paper is organized as follows. In Section II, we discuss the system model. In Section III, the simulation results are presented and, finally, conclusions are drawn in Section IV.

II. SYSTEM MODEL

In FDM uplink wireless transmission, the spectral efficiency can be improved by reducing the inter-carrier frequency spacing. Thereby, inter-channel interference increases. The ICI greatly depends on the normalized inter-carrier frequency spacing $\Delta_f T$, where T is the symbol interval and Δ_f is the frequency spacing in Hz between carriers. The carrier spacing $\Delta_f T$ is fixed to the value needed to achieve a desired Asymptotic Spectral Efficiency (ASE).

 ISBN: 978-1-61208-305-6

The ASE_{MIMO} of a frequency-division multiplexed uplink MIMO transmission with inter-carrier frequency spacing $\Delta_f T$ is defined as

$$ASE_{MIMO} \triangleq \lim_{E_b/N_0 \to \infty} SE_{MIMO} = M_T \cdot \frac{R_C \log_2 M}{\Delta_f T} \quad (1)$$

where R_C is the rate of the puncturedchannel code,M is the size of the CPM input alphabet and M_T is the number of transmit antennas, which is assumed to be $\leq M_R$, the number of receiver antennas. In [4], it was shown that it is possible to obtain a major improvement in SE thanks to a simple iterative ICI cancellation technique applied at the multiuser CPM receiver. The overall receiver complexity grows only linearly with the number of users. In this paper, MU FDM-CPM systems with $M_T x M_R$ antennas (MIMO($M_T x M_R$) *systems*) are investigated, their performance is assessed and compared to Single-Input Single-Output (SISO) scheme. The MIMO schemes are used in their full spatial multiplexing configuration. Figure 1 shows a block diagram of the proposed system. At the input, each *k*th user binary information sequence $a_0, \ldots, a_{k-1}$, is converted into several (M_T) parallel streams. Each data stream is conveyed to one of M_T encoding modulators, each consisting of an outer Convolutional Encoder (CE) connected to the inner CPM modulator through an interleaver. A rate 1/3 systematic recursive CE with four states and good distance properties [11] has been chosen. Its connection polynomials (in octal notation) are $(7; 5; 3)_8$, where 7_8 represents the coefficients of the feedback polynomial [9]. In order to achieve higher rates, the CE output is punctured as in [9]: a rate-matching algorithm is used to obtain coding rate R_C=3/8. The interleaver that connects the outer encoder to the CPM modulator is a symbol, spread (S-random) interleaver [12] whose parameters are set according to the code word size. The convolutional encoding, puncturing, interleaving, and modulation are realized by the CE-CPM blocks shown in Figure 1. Each user signal is characterized by a distinct phase φ_k and delay τ_k, as typically occurs in uplink systems. Ideal power control is considered here. As a result, the received signal power is equal for all users. The signal at the receiver input can be written as

$$\mathbf{r} = \sum_{k=0}^{K-1} \mathbf{H}_k \mathbf{x}_k + \mathbf{u} = \mathbf{H}\mathbf{x} + \mathbf{u}. \quad (2)$$

where $\mathbf{r} \in \mathbb{C}^{M_R}$ is the received signal vector, $\mathbf{H_k} \in \mathbb{C}^{M_R x M_T}$is the channel matrix of user k with elements representing the fading coefficients between the transmit and receive antennas, $\mathbf{x_k} \in \mathbb{C}^{M_T}$ is the transmitted symbol vector of user k, $\mathbf{u} \in \mathbb{C}^{M_R}$is the additive noise, modeled as a zero-mean complex Gaussian random vector. $\mathbf{H}$ is the joint channel matrix $M_R x k M_T$ and $\mathbf{x}$ is the joint transmitted symbol vector of length kM_T.

Each receive antenna receives a faded superposition of the M_T simultaneously transmitted signals corrupted by additive white Gaussian noise. The fading is assumed to be flat anddistributed according to a Rayleigh *pdf*. The random path gains between transmit antenna *i* and receive antenna *j*, $g_{i,j}(t)$, are independent complex Gaussian random variables with zero mean and variance per dimension 1/2. The fading is slow, such that the $M_T x M_R$fading coefficients are constant during a frame, but vary from frame to frame. The AWGN noise components $n_j(t)$, are independent zero-mean complex Gaussian random processes with power spectral density N_0. The received signal on antenna *j* is then:

$$r_j(t) = \sum_{k=0}^{K-1} \sum_{i=0}^{M_T-1} g_{k,i,j} x_{k,i}\left(t - \tau_k, \boldsymbol{a}_{k,i}\right) \quad (3)$$

$$\cdot e^{j(2\pi k \Delta_f t + \varphi_k)} + n_j(t), \quad j = 0, \ldots, M_R - 1$$

The MIMO receiver in the proposed FDM-CPM system (see Figure 1) uses a Minimum Mean Square Error (MMSE) multiuser detection technique [6] and low-complexity iterative algorithm to ICI cancellation. The MMSE block computes the cost function, i.e., minimizes:

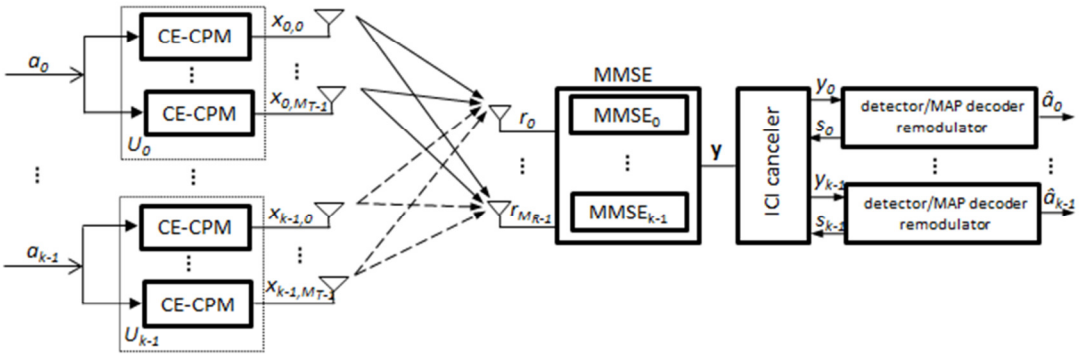


Figure 1. Block diagram of the MU coded FDM-CPM MIMO system.

 ISBN: 978-1-61208-305-6

$$E[(\mathbf{Mr}-\mathbf{x})(\mathbf{Mr}-\mathbf{x})^H]. \quad (4)$$

It amounts to finding the elements of matrix **M**:

$$\mathbf{M} = (\mathbf{H}^H\mathbf{H} + N_0\mathbf{I})^{-1}\mathbf{H}^H \quad (5)$$

where $\mathbf{H}^H$ denotes the conjugate transpose of matrix **H** and **I** is the identity matrix. Finally we obtain the signal vector **y** as

$$\mathbf{y} = \sum_{k=0}^{K-1} \mathbf{M}_k\mathbf{r}_k. \quad (6)$$

The function of the MMSE block is to compensate for the effect of the channel by inverting the channel matrix according to the MMSE criterion (4, 5). The signal **y** from MMSE reaches the ICI cancellation block. The receiver carries out ICI cancellation through a set of single-user MAP detector/remodulator blocks, as described by Perotti et al. [4]. The remodulators make use of the output of the MAP detector to compute the remodulated signal $s_k^{(i)}(t)$ relative to the *k*th user and *i*th iteration. The channel decoder performs two iterations loops. The *inner* loop is formed by the ICI canceller, the MAP detector, the CPE SISO decoder and the remodulator, while the *outer* loop involves the CPE SISO decoder, the CE SISO decoder, the interleaver and the deinterleaver between the inner CPE decoder and the outer CE decoder. ICI cancellation can be performed while executing the decoding iterations to enhance the receiver performance. In such case, after the inner CPE decoder is executed, remodulation is performed. Then, interference cancellation is performed and the CPM receiver, including the inner CPE decoder, is again executed. The decoder starts decoding a received code word executing N_{IC} *inner* iterations. Then, it executes N_D times an *outer* iteration followed by an inner iteration. This way, ICI cancellation is performed as part of the decoding iterations and it results in an improved ICI cancellation [4]. On the final *outer* iteration, a decision is made on the transmitted data symbols $\hat{a}_0, \ldots, \hat{a}_{k-1}$.

III. NUMERICAL SIMULATIONS AND DISCUSSION

Computer simulations have been made to evaluate the performance of the proposed MU serially concatenated CE FDM-CPM MIMO system. Different combinations of parameter setups for the MU FDM-CPM MIMO systems have been simulated. The most representative results have been selected for the presentation (hence the choice of parameters $\Delta_f T$, *M*, *L*, *h*). For comparison, performance of the MU FDM-CPM SISO system and the SISO and MIMO(2x2) systems with single user (no ICI) are also evaluated. We assume perfect knowledge of CSI at the receiver. In the considered system, we use full response (*L*=1) CPM modulation with the following parameters: *M*=4, *h*=1/5, *L*=1, RECtangular pulse shape (REC), $\Delta_f T = 0.5$, R_C=3/8, which implies an *ASE*=1.5 bits/s/Hz for the SISO system, and 3 or 4.5 bits/s/Hz for the MIMO system with two or three transmit antennas (M_T=2, 3), respectively. Simulations have been executed using information data words consisting of 1000 bits. The number of iterations in the receiver was experimentally fixed as a good trade-off between receiver performance and complexity. One ICI cancellation iteration *(*N_{IC}=4*)* is performed before decoding, then four decoding iterations (N_D=4) are performed.

The main results of this investigation are shown in Figure 2 and Figure 3. The spectral efficiency and bit error rate of the considered system are provided. Figure 2 shows the SE of the proposed FDM-CPM MIMO system. We observe that the MIMO transmission with ICI cancellation exhibits a significant SE improvement in comparison to SISO systems. Results show that spatial diversity can be exploited without the need of complex space-time coding techniques by using the proposed receiver, thus leading to considerably improved utilization of the available channel degrees of freedom comparing to the SISO case. Finally, we show the error rates obtained for proposed MU FDM-CPM MIMO receiver, Figure 3.

Different combinations of parameter setups for the MU FDM-CPM MIMO systems have been simulated.
Performance of the proposed receiver has been assessed and it has been shown that when the intercarrier frequency spacing, modulation scheme, and code rate are carefully chosen, performance close to the single-carrier (no ICI) may be obtained. The results for SISO and MIMO(2x2) systems show that for the BER=10^{-4} performance of the multiuser systems with ICI cancellation is close to that of the single-carrier systems. The curves (SISO) and MIMO(2x2) are as close as 0.5 dB and 0.7 dB to the BER curves (no ICI) for no ICI SISO and MIMO(2x2) systems, respectively.

The results in Figure 3 also show that the E_b/N_0 at BER=10^{-4} obtained by enhancing the number of receiver antennas from 2 to 4, while keeping constant the number of transmit antennas, in the proposed system, equals 8 dB. For BER=10^{-4} the MIMO(2x4) system yields performance improvement of about 3 dB with respect to the MIMO(3x4) system but the MIMO(2x4) system achieves lower *ASE* (3 bits/s/Hz) than the MIMO(3x4) system (4.5 bits/s/Hz). The same may be observed comparing the BER for the SISO system and the MIMO(2x2). In this case, the degradation is about 1 dB but the *ASE* for the MIMO(2x2) system is twice as large as in the SISO system. The degradation of the SISO system with respect to the MIMO(2x4) and MIMO(3x4) systems is about 7dB and 3.5 dB, respectively, with BER of about 10^{-4}. Additionally, the MIMO(2*x*4) and MIMO(3x4) systems prove higher *ASE* than the SISO system.

 ISBN: 978-1-61208-305-6

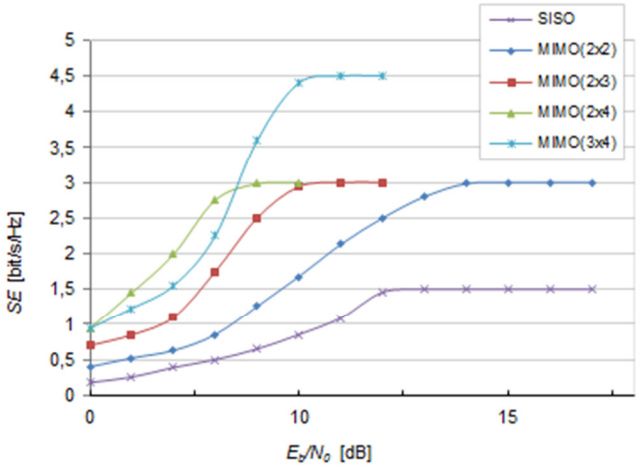


Figure 2. Spectral efficiency of MU coded FDM-CPM MIMO system with $\Delta_f T$=0.5, h=1/5, M=4, L=1, REC.

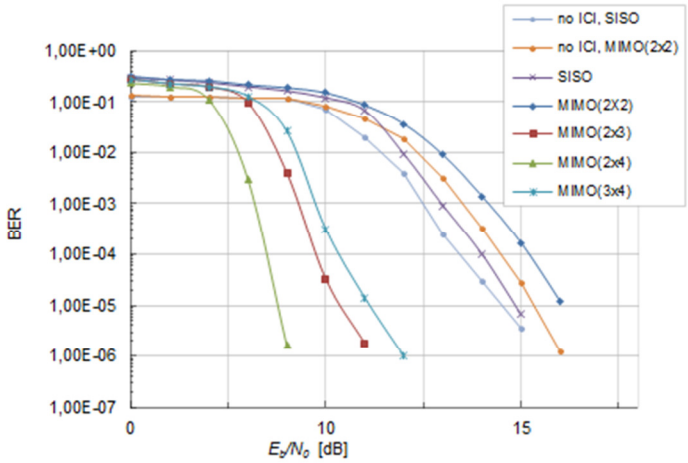


Figure 3. Bit error rate for the MU coded FDM-CPM MIMO system with various antenna configurations.

We compare the obtained results with those presented in [13], where MIMO CPM systems are designed. In [13], incoherent demodulator was adopted for full response MIMO CPM systems using blind signal separation in the receiver to separate the signals without any knowledge of the MIMO channel. The CPM scheme used therein is, e.g., quaternary with raised cosine (RC) transmit filter, h=1/4 and antenna configuration 2x4. Moreover, in [13] only a single user system (no ICI) is considered. The best results reported in [13] show that BER = 10^{-4} is achieved at E_b/N_0 close to 20 dB. Our MIMO(2x4) scheme with convolutional encoding reaches BER = 10^{-4} at E_b/N_0 =7 dB. The receiver in our system operates in presence of strong ICI and has perfect knowledge of the MIMO channel.

IV. Conclusion

In this paper, a multiuser coded FDM-CPM MIMO system has been proposed. Through simple MMSE-based multiuser detection and low-complexity iterative ICI cancellation, considerable improvements in both BER and SE are achieved with respect to single antenna systems, while the multiuser receiver complexity is kept low. The performance evaluation has been presented to demonstrate the superiority of the proposed multiuser FDM-CPM MIMO systems.

Acknowledgment

This work is supported in part by the Polish National Science Center under research grant 2011/01/B/ST7/06578.

References

[1] P. Moqvist, “Multiuser Serially Concatenated Continuous Phase Modulation,”PhD thesis, Chalmers University of Technology, Göteborg, Sweden, Dec. 2002.

[2] D. Bokolamulla and T. Aulin, “Multiuser detection for continuous phase modulation over Rayleigh fading channels,” IEEE Communications Letters, vol. 9, Oct. 2005, pp. 906-908.

[3] A. Piemontese, A. Graell iAmat and G. Colavolpe, “Frequency Packing and Multiuser Detection for CPMs: How to Improve the Spectral Efficiency of DVB-RCS2 Systems," IEEE Wireless Communications Letters, vol. 2, Feb. 2013, pp. 74-77.

[4] A. Perotti, S. Benedetto and P. Remlein, “Spectrally efficient multiuser CPM Systems,” IEEE International Conference on Communications, 2010, Cape Town, South Africa, May 2010, doi: 10.1109/ICC.2010.5501939

[5] E. Biglieri, A. R. Calderbank, A. G. Constantinides, A. Goldsmith and A. Paulraj, MIMO Wireless Communications. Cambridge University Press, 2010.

[6] X. Zhang and M. P. Fitz, “Space-Time Code Designwith Continuous Phase Modulation,” IEEE Journal on Selected Areas in Communications, vol. 21, June 2003, pp. 783–792.

[7] X. Zhang and M. P. Fitz, “Soft-Output Demodulation Space-Time-Coded Continuous Phase Modulation,” IEEE Transactions on Signal Processing, vol. 50, Oct. 2002, pp. 2589–2598.

[8] M. Hesse, J. Lebrun and L. Deneire, “Full Rate L2-Orthogonal Space-Time CPM for Three Antennas,” IEEE Globecom, 2008, pp. 3633-3637.

[9] R. L.Maw and D. P. Taylor, “Space-Time Coded Systems using Continuous Phase Modulation,” IEEE Transactions on Communications, vol. 55, Nov. 2007, pp. 2047-2051.

[10] P. Remlein, M. Jasinski and A. Perotti, “Receiver algorithm for coded multiuser CPM systems,” IET Electronics Letters, vol. 48, May 2012, pp. 633-635.

[11] A. Perotti, A. Tarable, S. Benedetto and G. Montorsi, “Capacity-achieving CPM schemes,” IEEE Transactions on Information Theory, vol. 56, Apr. 2010, pp. 1521-1541.

[12] D. Divsalar and S. Dolinar, “Weight Distributions for Turbo Codes Using Random andNonrandom Permutations,”JPL TDA Progress Report, vol. 42-122, April-Jun 1995, pp. 56-65.

[13] O. Weikert and U. Zolzer, “A wireless MIMO CPM system with blind signal separation for incoherent demodulation,” Advances in Radio Science, vol. 6, May 2008, pp. 101-105.

 ISBN: 978-1-61208-305-6

`

Wavelet Based Alternative Modulation Scheme Provides Better Reception with Fewer Errors and Good Security in Wireless Communication

Ramachadran Hariprakash
Arulmigu Meenakshi Amman College of Engineering,
Tamilnadu, India
rhp_27@ieee.org

Raju Balaji
Instrumentation Dept,
University of Madras
Tamilnadu, India
balajicisl@yahoo.com

Sabapathy Ananthi
Instrumentation Dept
University of Madras
Tamilnadu, India
ananthibabu@yahoo.com

Krishnaswami Padmanabhan
Anna University
Chennai
Tamilnadu, India
ck_padmanabhan@rediff.com

***Abstract* - In this paper, a different scheme of encoding digital data using wavelet functions instead of sinusoidal waves is explained. Data communications use various modes of encoding the data bits into a frequency signal. Phase shift keying of various forms such as Binary Phase Shift Keying, Quaternary Phase Shift Keying, etc., are in vogue in several current communication schemes, such as Global system for mobile communication. Errors in bits at the received end through a wireless channel are common in such data communication. These errors are mainly caused by improper phase changes in the detected audio signal. Since the method presented here with the use of wavelet functions provides several clues for identification of the data symbol instead of just by one criterion viz., the phase of the carrier as in the Phase shift keying schemes, this method is found to be better in performance. Simulation of the scheme was made with Matlab and the results provide an improved bit error ratio. Also, by varying the wavelets as per the user's choice, provides a higher level of security.**

Keywords - Data encoding; PSK Modem; Wavelet functions; Daubechie Wavelet; Bit Error Rate.

I. INTRODUCTION

Data communication has been employing several modes of encoding the data bits into an analog signal. Phase shift keying (PSK), in its various forms, such as Binary PSK, Quaternary Phase Shift Keying (QPSK), etc., use a baseband sine waveform with different phase positions to encode the data bits[1]. So far, alternatives to the sine wave signal for modulation have not been considered in any existing communication scheme, wired or wireless. All of the current methods employ a combination of phase shifting and amplitude changes for encoding the data bits. For example, in Figure 1, the QPSK modulation method uses four phases of the sine wave, with 90^{o} phase difference between symbols. In phase notation, the four waves can be represented as 0.7+0.7j, 0.7-0.7j,-0.7+0.7j, and -0.7-0.7j (Fig.2).

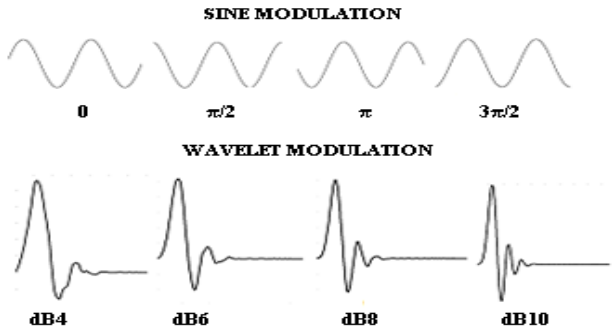


Fig. 1. Comparison between existing sine wave QPSK modulation and proposed wavelet modulation.

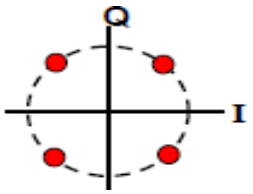


Fig. 2. The phase plot of QPSK symbols are having values in the complex plane marked.

When symbols of various values are thus encoded as analog sine modulation signals, we get a problem of spectral leakage, since there are discontinuities from symbol to symbol waveforms. This introduces errors in the case of transmission with multiple carriers. For instance, a truncated sine wave for symbol 01 will have a spectrum which spreads and is not confined to a single point. The shape of the spectrum will be somewhat similar, as shown in Figure 3.

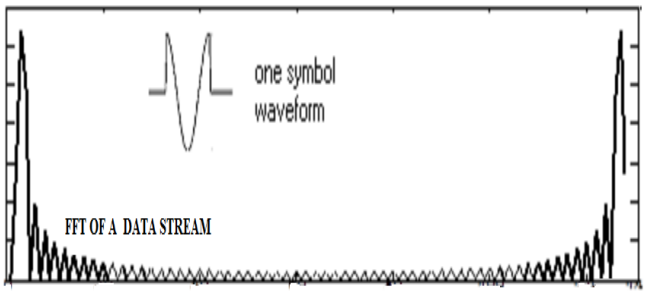


Fig. 3. The Fourier Transform (absolute value) of a data stream in a plain QPSK encoding modulating method is shown.

An alternative modulation scheme using waveforms of standard wavelet functions of the Daubechie (DB) type is shown in Figure 1. These functions are mathematically well defined and possess properties suitable for encoding and decoding. Four such wavelets, the DB4, DB6, DB8 and DB10 are shown as used for the same four symbols of data (00, 01, 10, and 11).

Trying such alternative modulation is done with a view to provide fewer errors in reception and also with some security provisions. This is done by using wavelets assigned to the bit patterns that can be of user centric.

Amongst the different wavelets known, such as Meyer [5], Coiflet [5], etc., the Daubechie wavelet [5], alone have several waveforms available for its different K-values and thus, it enables encoding more bit patterns per symbol.

To understand the genesis and properties of Daubechie wavelets, one can refer Addison [2]. During the symbol time, the end points of the waveform reach zero level and there is no discontinuity from symbol to symbol. This

 ISBN: 978-1-61208-305-6

`

helps in having a spectrum without any leakage. Well defined spectra result for such waveforms. So, in multiple carrier modulation methods as used in 802.11 and related schemes, the method can be applied with definitely much better results.

To point out the lacuna in an existing 802.11 multi carrier scheme, consider how the time signal is formed in that method, as clearly depicted in Figure 4. In this figure, for simplicity, five subcarriers are shown. Each subcarrier can have a particular phase angle of a few cycles of the sine wave used for modulation and this is represented as a phase with a real part and an imaginary part, as shown in the spectral representation as I and Q.

Each carrier is at a frequency range which is twice, thrice etc. of the first carrier. In order to generate the total time signal, which will include all the carrier frequencies, from the low to the highest, in it, the present method just *assumes* that the symbol waveform can be represented as a phase of single value, such as 0.707+0.707j for one such QPSK symbol. From Figure 3, we note that is incorrect, because of the discontinuous nature of the segment of a QPSK signal. In Global System for Mobile Communication (GSM), a Gaussian filter is applied to the symbol waveform called Gaussian minimum shift keying. Even still, there are discontinuities between symbols in the signal in time domain.

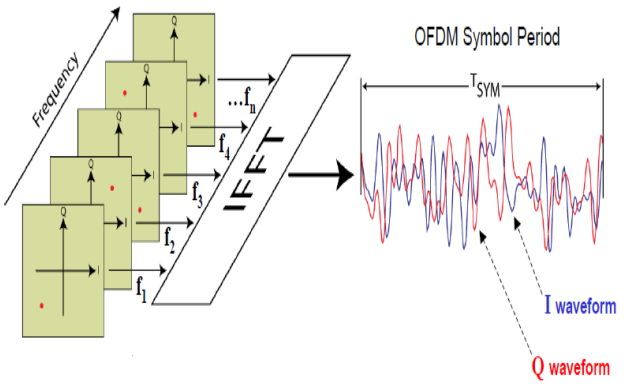


Fig. 4. The present method of forming a time signal for RF modulation in a Multi-carrier baseband scheme [3].

The mistake committed in the present scheme for getting the time signal from multiple QPSK (or even higher density schemes like 16QAM, 64QAM, etc.) lies in assuming that the spectrum of each carrier is a single point phase. Only if the sine wave is continuous for all time, the Fourier spectrum will be a single phase. So, combining these single phases at f1, f2, f3… etc. through an inverse Fourier Transform (FT) to generate the total time slot signal is having spectral leakage errors. These spectral leakage errors can definitely contribute to inter-symbol interference while decoding the signal at the received end.

The rest of the paper is arranged as follows. Section II gives a comparison of errors for standard QPSK and wavelet modulation. Section III deals with the bit error estimates for the proposed method. Then, the following Sections IV and V deal with wavelet shift keying multi-carrier communication as compared to sine phase shift modulation. The test implementation is discussed in Section VI and the comparison with sine modulation is detailed in Section VII. Section VIII deals with other attempts [12] bearing the name wavelets [13][14] and finally, the paper ends with a conclusion section.

II. ERRORS IN SYMBOLS – A COMPARISON

The process of decoding a standard sine modulated (QPSK, 16 QAM) signal first requires the generation of the reference sine wave from the first few data symbols, which are the initialization symbols. A Costas loop is a technique [4] that is also used to generate this reference wave. If the reference signal is slightly out of phase, that gives errors throughout the data.

Then, a multiplication of the signal received with this reference is performed. The average of this product is proportional to the cosine of the phase difference between the reference and the received segment of the phase shifted sine wave in that time slot. For each time slot, the phase is thus obtained. From this phase, the data symbol is decoded. With additional amplitude modulation as in QAM mode, the amplitude of the signal is also used to determine the symbol value.

In other words, the only information used for decoding is the phase of the signal. Phase of a sine wave is often subject to delays and changes en-route that is the reason for the errors in the received data.

Let us now describe the decoding method when wavelet signals are used. A wavelet signal is a well defined mathematical signal, which has a compact support and a finite spectrum without spreading like the phase shifted sine wave segments. The pattern of each of the Daubechie signals is different and specific [5]. Thus, there are several criteria available for determining, in any one time slot, which wavelet is received.

First of all, from the received data, which is converted into digital numbers from the analog to digital convertor in the receiver, we have to isolate the symbols. This is the process of synchronization. In this case, it is much easier than for sine phase modulated signals. The first few data symbols are known and are identical. The first peak and then the peak of the second symbol are fetched. The midpoint of data is a starting point for the second symbol. We know the number of data points in a symbol from the two peak positions. From then on, the data samples can be isolated easily.

The wavelet functions possess several peaks. The following criteria for decoding have been used by the authors.

i). RMS value of the autocorrelation function:

This calculation resembles the calculation made in the sine modulation existing methods. Correlation is multiplication, shifting and addition, and the estimation of the Root Mean Square (RMS) value.

ii). Peaks their ratios and spacing is another criterion:

Among the DB4 to DB20 wavelets, the peaks differ. The number of peaks increases for the higher degree wavelets. To determine the peaks is a simple calculation. By comparing with the values of the received data in each time slot, the correct symbol value is found.

 ISBN: 978-1-61208-305-6

`

A check on the results of the above two methods is able to infer plausible symbol errors.

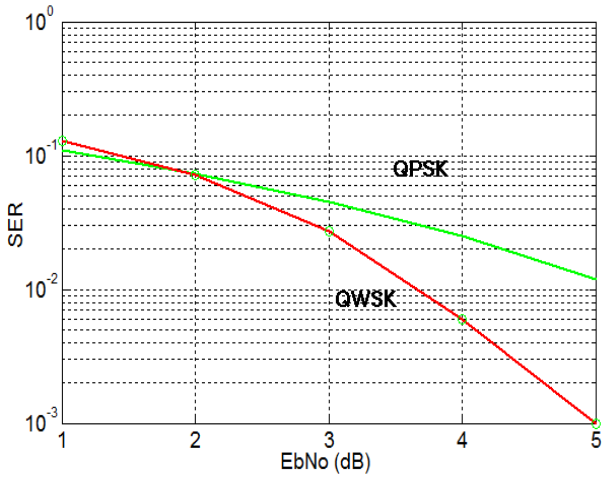


Fig. 5. Errors simulated by Monte Carlo simulation for the QPSK and our wavelet modulation schemes for different signal to noise ratios.

We conducted the tests with 16 wavelets representing a four bit data symbol and obtained good results.

For example, the Matlab sample demo program for QPSK [6] sine phase shift modulation scheme gives error rates more than our 4 bit wavelet based keying method, as shown in the graph of Figure 5.We note the good reduction in errors with the method.

III. BIT ERROR ESTIMATE COMPARED TO PLAIN QPSK

In sine modulated QPSK transmission, the bit error is estimated as the probability of finding the phase position correctly. This is based on the conditional probabilities of two vector positions [7], i.e., the two axes of the phase. The probability of error per symbol is denoted by P_e and is given by

$$P_e = \text{erfc}\ (\sqrt{E/N_0}) \qquad (1)$$

Where a white Gaussian noise (variance N_o /2) is assumed to affect the transmitted signal. This result was obtained by integrating up to infinity, since the vector can take a position all along the axis of the phase. Such a result has also been verified by Matlab's simulations using the Monte Carlo method [6].

In this WSK method, the probability of getting a wrong wavelet symbol instead of the correct symbol at any instant depends on the probability of nearness to the right symbol. The root mean square criterion of the autocorrelation function of the received symbol has the values, which decrease from DB4 to DB32, monotonically. These values are stored by *a priori* calculations.

The probability of the value deviating far from the right symbol's value depends on the noise. The integral of the probability function of this Gaussian random variable is a definite integral between the two fixed values only and not from 0 to infinity as in the derivation of the QPSK bit error probability.

TABLE I. COMPARISON TABLE BETWEEN 16 QAM AND 16 WSK

	Symbol synchronisation	***Detecting the Symbol***	***Symbol assignment***	***Bit Errors by Simulation***
16 QAM	Synchronisation requires generating a reference sine wave, which is difficult and if there are slight errors in the reference, that will reflect all data bits.	This symbol detection is based on both amplitude and phase position of the sine waves within the time slot. With very few cycles sent in each slot, the phase detection is often erroneous due to the analog signal variations at the edges of the symbols.	Symbols are assigned values so that nearby constellation points (in phase diagram) differ only in one bit value. That constrains the values of the symbols with respect to the phase shift values.	By a simulation it is possible to send around 10000 symbols randomly and add Gaussian White noise of a definite Signal to noise ratio. Then by decoding, the data can be compared and symbol errors counted. The error rate is higher than for the WSK method.
16 WSK	Synchronisation is easy after finding the first two data symbol's peak position and from then on, based on fixed time sampled allotted to each symbol.	Here, the detection is based on more than one criterion. The first and best criterion is (1) above. By using more criterion and comparison, confidence levels of symbols can also be found.	Here, there is no constraint on bit values & encoding wavelet numbers. We can shuffle the values of wavelet nos. with respect to the symbol values, which can be used for encryption.	Simulation similarly by a large number of data symbols encoded, noise added and decoded gave better results.

Thus, the value is definitely less than in the case of the sine modulated method. The errors stimulated by Matlab for Monte-Carlo simulation for QPSK and the wavelet modulation scheme is shown in Figure 5. Table I above shows the comparison table regarding the features like the symbol synchronization, detection of symbol, etc., between the 16QAM and 16WSK.

IV. PROPOSED WAVELET BASED SCHEME SIMILAR TO OFDM IN SINE PHASE SHIFT MODULATION

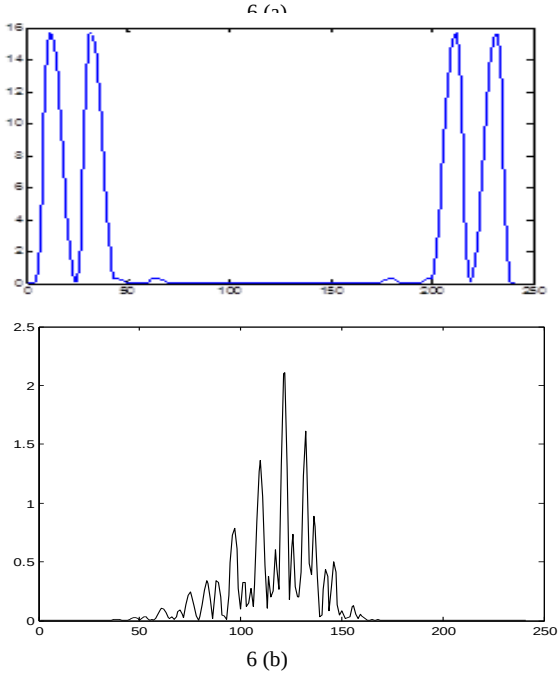
In the 802.11 a-g WLAN example, there are 48 subcarriers. The symbol time is 4 µs. The carrier spacing is 312.5 KHz [8].

Similar to this, if we want to implement the wavelet based keying scheme, let us examine the details of its implementation in what follows. For each time slot, there are many subcarriers with a spacing of frequency between them. We have to generate the time signal for all such

 ISBN: 978-1-61208-305-6

`

carriers put together. As the example from Figure 6a shows, addition of a second carrier with a DB8 waveform to the first carrier will mean adding a shifted version of the spectrum of the DB8 signal to the total spectrum. Thus, there are two spectral peaks shown for two carriers numbered 1 and 2 in Figure 6a. The time signal for this will be as in Figure 6b. This is obtained by the inverse Fourier transform.



6 (b)

Fig. 6 (a). An addition of a second sub carrier as a second spectrum added with a shift of frequency of 1/8 of the sampling frequency.

Fig. 6(b). The time signal obtained by using the above two sub carrier modulations. Only 2 subcarriers are shown for clarity.

For example, let us take for simplicity an eight carrier system. There will be eight symbols sent in one time slot. These eight symbols will be formed by the data at the current time slot. Let the data for these 8 carriers are, say, DB4, DB8, DB12 …… DB20. We have pre-calculated spectra for each of the several encoding wavelets. We just position these spectra at the frequency slots 1 to 8 in this order. For Figure 6a, it was done for just two carriers. For three carriers, Figure 7a illustrates the signals for each of the carriers. The process of positioning the spectra peaks and arriving at the overall spectrum at any one time slot is a simple operation. The Inverse Fast Fourier Transform (IFFT) of the total set of such spectrum will give the total time signal.

The points in the Inverse Discrete Wavelet Transform (IDWT) will be just the product of the number of subcarriers and the number of data points in one symbol. In Figure 7, 128 point IFFT space is chosen. This gives 64 points upto the folding frequency. If 8 subcarriers are used, each subcarrier space is 8 points in the total of 64, being half the total IFFT space. When combining the data for each subcarrier, we have to shift the spectrum of the corresponding wavelet for its symbol. Thus, the positioning of the pre-stored spectra for all the 16 wavelets (in a scheme similar to 16QAM) can be done to form the spectrum of the total time signal. This is inverted and it is a 128 point IFFT. This signal is modulated with the RF and transmitted.

Instead of this complicated method involving time for IFFT calculation, the following method is adopted here.

We know the wavelets used for encoding and also the subcarrier frequencies. The received RF demodulated signal would be a time signal comprising of all these wavelets. To retrieve each of the subcarrier data, we do an FFT of that signal. That FFT signal has to be separated into 8 different spectra by splitting the same. That requires a splitting program. Each of the split spectra is inverted to yield the DB waveforms for each subcarrier symbol. The illustration in Fig. 7b shows how the waveforms for three subcarriers merge very well with the original waveforms for the symbol. Thus, the symbols could be retrieved from all the subcarriers.

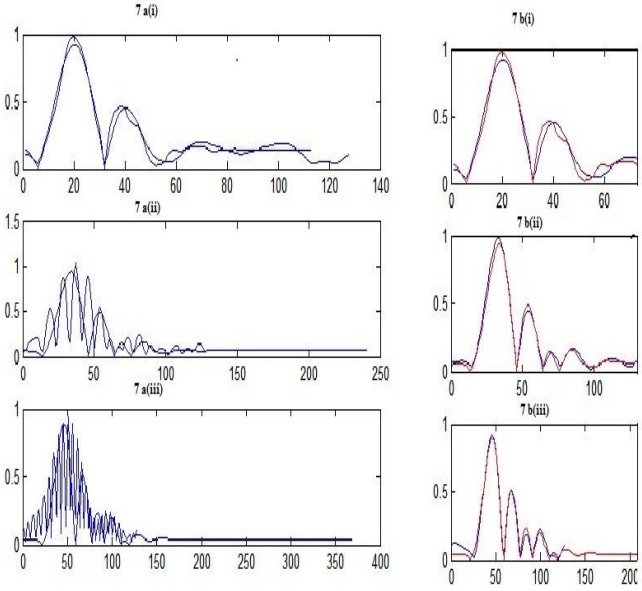


Fig. 7 a i). The DB4 signal in the first time slot, superimposed with the signal created by a truncated IFFT. 7a ii) The DB8 signal if it is shifted one subcarrier right then shows higher frequency waves in it after the IFFT. 7a iii) The DB12 signal shifted two subcarriers right then it indicates still higher frequencies.

Fig.7 b (i), 7b(ii), 7b(iii) Shows the respective received signal after frequency separation, the received data merges with the symbol data perfectly.

The above Figure 7 just illustrates the method, though, in practice, the data is actually collected from more subcarriers and then decoded. The above simulation is just to illustrate that inter carrier interference is totally absent here. For sine modulation schemes, the OFDM simulation is found in [9].

V. MULTIPATH ERROR ESTIMATION

Whenever multiple carriers are used, multipath reception always leads to errors more than desirable. The estimate of multipath errors for a general sine modulated multicarrier signal is discussed elsewhere [10]. To make a comparison of the OFDM or QAM schemes with this proposed WSK scheme, let us examine the effect of a second path signal adding to the direct path. Let us assume that, as usual, the second path arrives at the next

 ISBN: 978-1-61208-305-6

`

time slot and hence, in the second timeslot, we have a reception of the combination of two symbol signals.

Let us compare the performance with respect to inter-symbol interference for the sine and wavelet based schemes.

i) Sine phase modulation method:

Let us consider that the second (longer) path signal is attenuated by about 50% to that of the directpath signal and that the phase of that symbol is likely to be any one of the constellation points in phase space. If the two phases are very adjacent ones, then the error in phase by combining the two path signals will be less than half the phase difference between these two phases. But, if the previous symbol is a far apart symbol in phase diagram, say 135^o away from the current signal, then the addition of the two phases moves the net phase by as much as 45^0 out of correct position (see Fig.8).

ii) Wavelet based MWSK scheme:

In the proposed WSK method, the two symbols could be say DB8 and DB4 on the direct and secondary paths, which when added gives a composite signal. With a 50% of the latter signal which is a wrong signal, the net root mean square values of the two will be obtained by squaring and summing and again taking the root.

Thus, for example if 352 for DB4 and 115 for DB8 are used for this Fig.8, we get the net Fig. as

$$\text{Rms} = \sqrt{115^2 + (352/2)^2} = 210 \qquad (2)$$

which is nearer to the DB8 and hence it still yields the correct data. It can be similarly shown in all cases, with even 50% indirect path signal, the correct results are obtainable, which is one of the merits of the proposed scheme. This compares favorably with the sine modulated scheme.

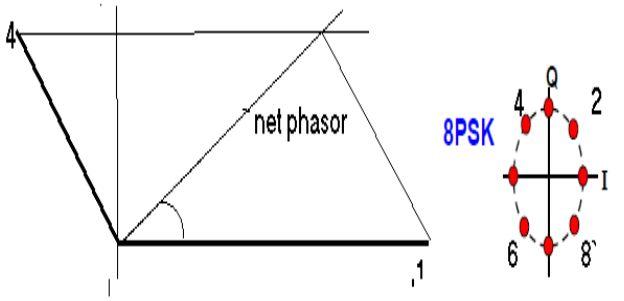


Fig. 8. In an 8 Phase Shift Keying sine modulation scheme, the path 2 signal belongs to no.4 and hence is at 135^o to the horizontal.

The signal on the direct path is a no.1 signal, which is horizontal and is the true signal. Since the combination of even a 50% attenuated path 2 signals (note the inclined phase is half in length) would shift the net phase by 45^o as shown by the parallelogram diagonal, the received symbol suffers an error. The comparison features of the proposed wavelet based scheme with sine based modulation scheme for multicarriers are given below in Table II.

TABLE II. COMPARISON TABLE BETWEEN SINE MODULATED O.F.D.M. AND WAVELET SIGNALS ON MULTICARRIERS

	Formation of the total time signal in each time slot	***Inter Carrier Symbol Interference***	***Multiple path and fading***	***Power levels***
OFDM (SINE)	This finds by table look up, the phase for each subcarrier's symbol based on the data. Then, it merges the phases in the frequency space as shown in Figure 4. Then, it generates the total time signal by an IFFT.	The spectrum of every subcarrier extends very much into the subcarrier previous and next to it. Because of the orthogonal property of sine waves, the multiplication with the correct subcarrier yields the phase value of the symbol. However, problems arise when waveforms are distorted or saturated. Wrong symbols are obtained.	Multiple radio frequency paths, such as reflections give added delayed signals to the receiver. These causes inter symbol interference. That is why, the method is not much usable for long distance wireless communication.	The power level in the modulation is sum of the subcarrier powers based on the RMS value of the sine wave voltage value used for modulation. The modulating power is more in this case than for the wavelet multicarrier.
MULTI CARRIER (WAVELET)	Here, we have samples of the several wavelets' signals are stored already in the different scales of frequencies. We just add the signals corresponding to each subcarrier's encoded wavelet.	In our method, there is no merging of the subcarrier frequency spectra as seen from Figure 6a. Thus, subcarrier signals are separable without any mixing. Hence, the performance is likely to be better.	Here, the addition of a reduced amplitude signal with delay from a reflected path will not affect the calculation involved in deciding the symbol. We have multiple criteria for decoding and hence the decoding is more definite and with plausible errors, confidence levels can be obtained by combining the results of the multiple decoding criteria.	The power level of the signal waveform from a wavelet used for encoding is less than that of a continuous sine wave. Thus, we have less modulating power in the scheme.

The method of combining the multiple subcarriers is through a look up table data, after encoding the symbols as wavelets for these multiple symbols.

 ISBN: 978-1-61208-305-6

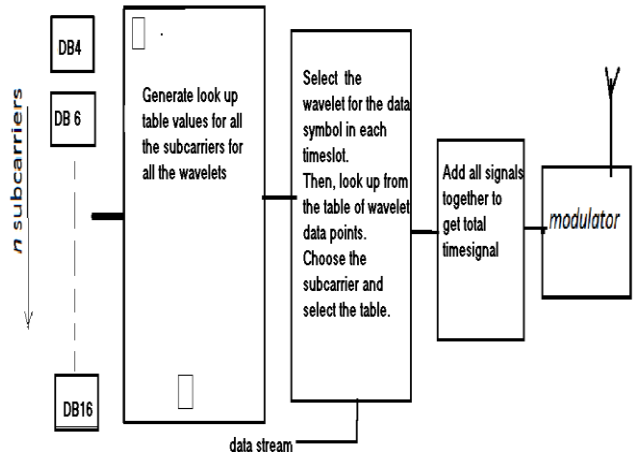


Fig. 9. Showing the scheme for multi carrier signal generation and modulation prior to transmission

The waveforms of the time signals are pre-stored and hence just addition of the waveforms will yield the composite waveform in time. This is shown in Figure 9, and is given to the RF Modulator.

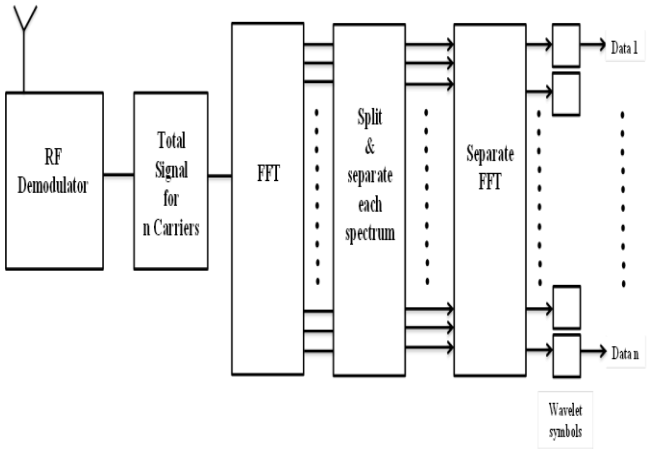


Fig. 10. Showing the method of decoding multiple carrier signals and separating the data symbols.

At the receiving end, the total demodulated time signal is transformed in Fourier space and the different symbol subcarriers are separated in this space. Each of them is inverse transformed to get back the wavelet waveform. By comparing the wavelet received with the encoding wavelets (Fig. 10), the symbol values are separately found.

VI. TEST IMPLEMENTATION

From the Matlab program, which generates the waveform of the time signal, we considered the transmission of the analog data through a RF signal generator with amplitude modulation facility. The signal is generated from the sound card of the computer using WINSOUND command on the Matlab. The sampling frequency is low, but is enough for testing; the sampling frequency value is 44.1 KHz.Therefore, we could send only 4 subcarriers fs = 44100 and sound sc (signal, fs) are the commands.

The signal from the sound card audio jack is connected to the modulator of the RF generator. The RF frequency is set to a radio frequency in the near short wave.

The receiver is tuned to this frequency, which is a communications receiver. The received data is again fed through the line input of the sound card of the PC. The PC reads the sound card audio using the command Analog input (AI) and other related win-sound commands. The audio signals are stored in user specified files. These signals are processed and the recognition of the wavelet is made, thus the data is created.

Digital data for wireless communication have to be converted to analog signals for modulation over a RF frequency for transmission. The method of encoding data bits has all along been using merely the phase shift and amplitude of a baseband (low frequency) sine wave. Alternative to the sine wave no other waveform has so far been tried out.

In our paper, we present the use of wavelets for modulation. We choose a particular type of wavelet for each data symbol (Table III). With 16 wavelet signals, we can encode four bits of data.

TABLE III. WAVELET SIGNALS AND THIER ENCODING

S.No	Bits	Wavelet
1	0000	DB 4
2	0001	DB 5
3	0010	DB 6
4	0011	DB 7
5	0100	DB 8
6	0101	DB 9
7	0110	DB 10
8	0111	DB 11
9	1000	DB 12
10	1001	DB 13
11	1010	DB 14
12	1011	DB 15
13	1100	DB 16
14	1101	DB 17
15	1110	DB 18
16	1111	DB 19

The wavelet signals as sent through a typical transmission test look like in Figure 11.The typical waveform for a short stretch of eight symbols is shown as transmitted and after noise addition.

 ISBN: 978-1-61208-305-6

`

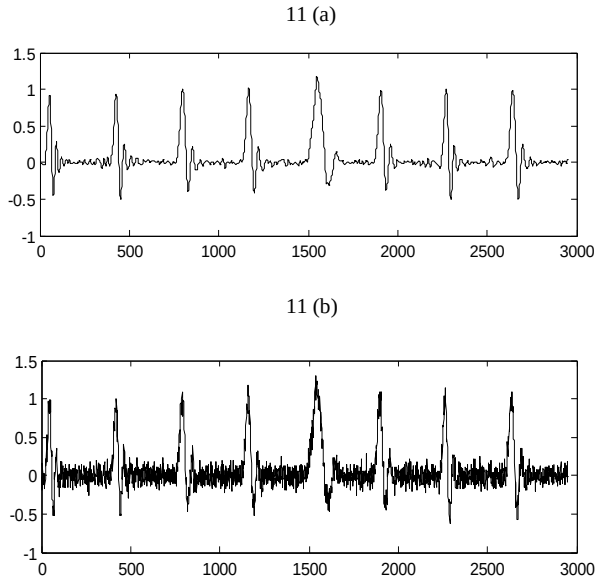


Fig. 11 a) Modulating wavelet signals for a set of 8 symbols.
b) The addition of white noise is seen.

With multiple carriers in a single time slot, we can send multiple symbols resulting in greater throughput.

VII. COMPARISON WITH SINE MODULATION SCHEME

Thus, wavelet based modulation has been shown to provide a better symbol error and also reduces multi path distortion compared to the existing sine modulation method.

Secondly, we consider how a security aspect can be included in the scheme. For this, let us consider how the standard QAM modulation scheme encodes the data bit values.

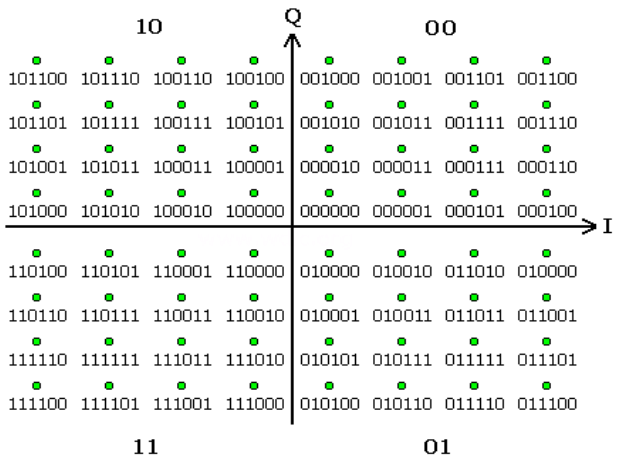


Fig. 12. Bit patterns used in the conventional 64QAM Modulation.

For example, the bit patterns versus the phase positions are indicated in Figure 12. It is noted that the data bit patterns have fixed positions in the constellation with a view to make only one bit change between adjacent phase points.

But, with our wavelet based functions, we have no such restriction. The wavelets assigned to the bit patterns can be the user's choice at any time and thus, by varying the pattern versus wavelet function table from time to time, an additional level of security is obtained [11].

VIII. OTHER ATTEMPTS BEARING THE NAME WAVELETS

From Figure 5, it is noted that the signal in time course is evaluated by an IFFT at transmitter; the FFT is done at receiver. Looking at the transform pair (FFT-IFFT), some authors have thought of replacing this by other known transforms. Immediately it suggests the ubiquitous wavelet transform. The ordinary wavelet transform however reduces the number of points and so the wavelet packet transform comes to mind. This principle is like the analogy of traveler carrying cash from country to another. If he goes to Europe, he can carry Euros, pounds, dollars or even some material of value. But, other than the Euro, all the rest suffer losses in transfer. Thus, in the WPT imagination, the difference coefficients are always small in value compared to the approximation coefficients and the small values get easily masked by even a limited amount of noise. The reconstitution using the approximation and differences will yield considerably large errors. The authors propounding such methods could only treat their concept at theoretical and partial simulation level and not even a baseband actual transmit receiver session could be reported by them. In the proposed method, the multi carrier scheme uses merely the Fourier space and the time signals are normal. Only in the symbol level, we use the waveforms of the scaling functions of the DB wavelets. The scenario of our scheme is useful at single carrier for cellular communication and with multi carrier communication for short range wireless in-house and wired LAN.

Most literature cites the use of Daubechie wavelets for reasons already mentioned. The other wavelets may not provide for our use as much as sixteen different patterns for encoding. Hence, the proposed work rests mainly on the DB wavelet scaling function waveform modulation. So, comparing with other possible wavelets was felt unnecessary.

In the paper by Matthieu Gautier, Marylin Arndt and Joch Lienard [12], the signal is viewed as a sum of modulated wavelet packets. They suggest using the IWPT and WPT pair at transmitter and receptor. The concept is dealt with mathematically and so is their simulation. But, the details of an implementable scheme are left out in so far as actual waveforms for encoding a message is concerned and no techniques as to decode the signal at reception are given. The work [13] is also a very similar attempt. In another paper [14] it describes the possibility of the same WPT transforms for multicarrier communication by the similar WPT reception but they do not give of any decoding of neither symbols nor do they deal with how many bits are encoded in a symbol and in what manner the signal based on the DWT is generated. This paper is a conceptual account with more details of the wavelets filter functions and spectral overlaps of the multiple carriers. The bit error curves given are based only on their assumed theoretical Gaussian error formulas.

 ISBN: 978-1-61208-305-6

`

IX. CONCLUSION

Data communication in a security system, separately set up for a private or similar requirement, can exploit the advantage of such a different modulation scheme over conventional PSK based schemes. The encoding scheme method using 4, 8 or 16 wavelets in schemes with 2 bits, 3 bits, or 4 bits encoding in a symbol can be selected by the user and the assignment of bit patterns for the data symbol can also be the user's choice and can be varied from time to time to provide a level of security.

The usage scenario of the scheme can be anything, such as in-house, short range wireless or cellular wireless. The plain wavelet encoding without multiple subcarriers will suffice as is employed presently in wireless cellular systems. In short range wireless, as in 802.11 schemes, the multiple subcarrier wavelet schemes is applicable equally well as the present sine modulated scheme.

Additional bonus in the scheme is the better decoding possibilities leading to fewer errors, as seen in Figure 5, even in the case of basic QPSK verus 4-WSK scheme.

REFERENCES

[1] L. Frenzel, "Understanding Modern Digital Modulation Techniques," Electronic Design Magazine, 23 Jan. 2012.[online].Available.http://electronicdesign.com/communications/understanding-modern-digital-modulation-techniques.

[2] P. S. Addison, "The Illustrated Wavelet Transform Handbook; Introductory Theory and Applications in Science, Engineering, Medicine and Finance," IOP Press, 2002, pp. 104-116.

[3] Keithley Instruments Inc., "An Introduction to Orthogonal Frequency Division Multiplex Technology," ©copyright2004[online].Available:http://www.ieee.li/pdf/viewgraphs/introduction_orthogonal_frequency_division_multiplex.pdf.

[4] S. Haykin, "Digital Communications," Wiley India, 2013.

[5] I. Daubechies, "Ten Lectures on Wavelets," vol. 61 of CBMS-NSF regional conference series in Applied Mathematics SIAM, Philadelphia, PA, 1992.

[6] B.P. Paris, "Simulation of Wireless Communication Systems using Matlab," Dept. Electrical and Computer Engineering, George Mason University, Fall 2007. [Online].Available.ftp://doc.nit.ac.ir/cee/m.zahabi/Courses/Wireless3892/simulation.pdf.

[7] J.G. Proakis, "Digital Communications," McGraw Hill Inc., New York, 4th Edition, 2001.

[8] K. Roebuck, "IEEE 802.11ac- High Impact Technology- what you need to know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors," Emereo Pty Limited, 2011.

[9] N. Pathak, "OFDM Simulation Using Matlab," International Journal of Engineering Research and Technology, vol. 1 (6), August 2012, pp. 1-6.

[10] J.P. Linnartz, "Multi-Path Scatter Function" – JPLs Wireless Communication reference website, copyright @ 1996.

[11] R. Hariprakash, S. Ananthi, and K. Padmanabhan, "Secured image using wavelets for spread spectrum communication in a remote surveillance system," Proc. Third International Conference on Advances in Human – Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC 2010), Dec 22-27, 2010, pp. 61-66, doi: 10.1109/CENTRIC 2010.15.

[12] M. Gautier, M. Arndt, and J. Lienard, "Efficient Wavelet Packet Modulation for Wireless Communication," Proc. Third Advanced International Conference on Telecommunications (AICT 2007), May 2007, p. 19, doi: 10.1109 /AICT.2007.21.

[13] A. Jamin and P. Mahonen, "Wavelet Packet Modulation for Wireless Communications," Wireless communications and Mobile Computing Journal, vol. 5 (2), March 2005, pp. 1-18.

[14] N. Nikolov and Z. Nikolov, "A Communication System with Wavelet Packet Division Multiplexing in an Environment of White Gaussian Noise and Narrow-Band Interference," Cybernetics and Information Technologies, vol. 5(1), Sofia 2005, pp. 100-114.

 ISBN: 978-1-61208-305-6

Towards a Dynamic QoS Management Solution for Mobile Networks based on GNU/Linux Systems

Adapting Already Existing Solutions to Vehicular Environments

Gorka Urquiola, Asier Perallos, Itziar Salaberria, Roberto Carballedo
Deusto Institute of Technology (DeustoTech)
University of Deusto
Bilbao, Spain
{gurquiola, perallos, itziar.salaberria, roberto.carballedo}@deusto.es

***Abstract*—The number of applications used in Intelligent Transportation Systems is growing very quickly. This implies a greater consumption of vehicular network bandwidth hence there could be a high probability of delay of priority requests in this networks. Consequently, an exhaustive control of the bandwidth is needed to provide a Quality of Service according to the demands of certain applications. In this paper, a communications middleware to provide the management of the Quality of Service and prioritize applications' requests on mobile networks is tested. The proposed system, in order to reduce development efforts, has been addressed only reusing and configuring already implemented and tested GNU/Linux based software utilities, originally designed to be used in non-mobile environments.**

Keywords-Vehicle-to-Ground Communications; Quality of Service; Requests Priorization; Virtual Private Network; Queue Disciplines; Linux

I. INTRODUCTION

In transportation systems, is not easy to guarantee continuous communications and a stable available network bandwidth inside the vehicles. Common network configurations used in non-mobile environments, such as the ones used in an office (where static network links are used, continuous communication can be assured using wired and backup links [1], and the service failure probability depends on rare environmental factors and internet service provider quality), are not directly adopted in mobile networks. The reason is that the quality of the communication could be affected by several dynamic factors, such as coverage changes according to the location of the vehicle, data packets losses or event cuts in the communication that may occur.

For such networks, it is usual to adopt vehicle to ground architectures [2], in which it is necessary to maintain the communication between the mobiles and control centre nodes or even the communication between all the mobile nodes.

Moreover, the number of applications used in this kind of mobile environment is growing in an exponential way due to the requirements of the Intelligent Transportation Systems (ITS) [3]. The mobile services offered by the internet service providers are not always capable of providing a suitable bandwidth that meets the needs of such applications. Consequently, an exhaustive control of the bandwidth consumption is needed to provide the Quality of Service (QoS) demanded by certain applications, such as in the case of surveillance video streaming (high bandwidth consumption, low priority) and an alarm trigger (low bandwidth consumption, high priority). In these cases, communication requests must be prioritized or delayed assuring priority to the most relevant data traffic and leaving in background the not critical one [4].

In order to have a greater connectivity and coverage, we could use 3G modems for accessing to the Internet. Instead of developing specific software able to manage the different links, establishing the active channel to use (based on factors such as coverage, availability and bandwidth), we decided to combine existing software tools. The aim is to get an easy to develop and deploy communication solution for mobile (vehicular) environments which is able to manage the QoS of applications in a dynamic way [5].

To achieve this target, a system based on a GNU/Linux distribution, using only free and open-source software tools, has been designed and tested. These software tools have a fairly widespread use, which incurs in having an always updated and well documented system. Thus, we can develop a communication system with a minimum initial investment and whose robustness and fault tolerance is guaranteed by the support and contribution of a community of worldwide developers.

The rest of the paper is organized as follows. In Section II, a brief overview of the state of the art is included. The contributions of the developed communication system are included in Section III. The proposed solution design, including the description of the tools used and the reasons for choosing them, is presented in Section IV. Then, in Section V, the real scenario in which the system has been tested is described. Finally, the results of the tests are analyzed in Section VI and the paper ends with the conclusions and future work.

II. STATE OF THE ART

Transportation companies demand greater efficiency for their systems, therefore, wireless communication technologies are growing in vehicular systems. Also, they are

 ISBN: 978-1-61208-305-6

also seeking to provide new information services [6]. For many years, the networks used in transport systems have been formed based on separate islands of physical media and protocols [7]. Currently, the existence of multiple transmission alternatives provides higher communication bandwidths [8], but this does not mean a better performance in regard to interoperability, temporary or reliability properties [9]. Thus, there is an increasing complexity in telematics contexts because of the continued growth of the specific systems and solutions, requiring technologies that enable greater interoperability between these solutions [10].

On the other hand, even if the emphasis in developing wireless networks is on network bandwidth and coverage, the applicability of the communication system will largely depend on their ability to provide sufficient data rates (QoS requirements), considering introduced protocol overhead, packet fragmentation and possible retransmissions.

Therefore, wireless communications applied to mobile environment present several limitations related to coverage and bandwidth that can cause service disruptions. Moreover, wireless stations that need to transmit critical information must deal with wireless stations wishing to transmit less priority traffic.

For the purpose of achieving QoS requirements demanded by services, several communication management and prioritization heuristics [11,12] and mechanisms exist [13-15]. Although existing solutions are mainly focused on network aspects and not in final applications and services, other approaches are focused on optimizing the use of the network technologies according to the type of traffic generated by applications (QoS control). Therefore, there is an open research field that can be tackled from two complementary points of view: (1) QoS requirements management, which involves technology concepts related to the information to transmit, and (2) aspects about network conditions that make possible the transmission of that information (bandwidth, coverage, latency, etc.). The work presented in this paper will explore this first approach.

There are multiple works regarding communications optimization, including traffic prioritization and QoS control. However, these works are usually focused on networks instead of applications or services that use these networks [6,16]. In addition, there are industrial solutions designed to respond to these detected communications needs and challenges in transportation systems [17, 18]. But, neither of these projects establishes a communication system that prioritizes data transmissions dynamically.

III. TECHNICAL CONTRIBUTIONS

There are three main technical contributions of the proposed vehicular communication system:

- *Network traffic regulation*. The number of applications used in vehicular environments is growing very quickly, which implies a greater consumption of network bandwidth. The regulation of applications network traffic becomes important, as an excessive network bandwidth consumption by a secondary application may cause delays or even data packets losses of a priority application. Thus, this can be a problem for those applications that consume lower bandwidth, but which have a higher transmission priority.
- *Network security*. Security in communications is another aspect to consider as applications may be required to transmit sensitive information between the mobile node and ground centres, such as ticketing information using Near Field Communication (NFC) or contact/contactless SmartCards.
- *Onboard subnet management*. It would be desirable that the onboard communication system had to be able to manage a subnet and serve as a gateway to the control applications located on ground centres. It is a requirement that does not require to run all applications on the same device, allowing the use of additional devices, such as sensors or IP cameras, which act as additional nodes in a subnet.

IV. COMMUNICATIONS SYSTEM DESIGN

The proposed system and the later tests have been addressed in a generic manner, as an assumption of the authors, trying to cover a wide range of use cases. Thus, they do not represent real expectations of specific manufacturers and users. Nevertheless, a specific use case of the proposed solution is train-to-earth railway communications [17]. The train units need to transmit heterogeneous information (different size and urgency). Thus, for example, a train requires that critical positioning data of few kilobytes to be transmitted continuously, while other type of data transmission like video streaming may be heavier but can wait to be transmitted until priority data has been sent. The management of these kinds of communications requires very different priority and QoS treatments that could be addressed by our work.

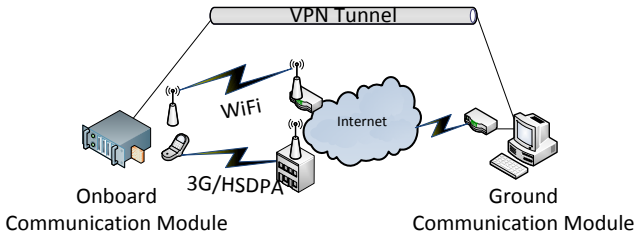


Figure 1. Conceptual architecture of the system.

The proposed communications system follows a vehicle-to-ground architecture based on the existence of an Onboard Communication Module (OCM) and a Ground Communication Module (GCM). The onboard module has two different wireless communication links – a 3G modem and a WiFi antenna – and the ground module has a wired broadband link (Figure 1). The onboard module will only use one of the available links (we refer to it as the active channel). The system will use the WiFi link if a known access point is available, if not, the system will use the 3G modem as network link. A 4G/LTE connection would be another way to implement the mobile connection, but

 ISBN: 978-1-61208-305-6

because of the ease of implantation in a wide range of scenarios the 3G option has been finally chosen for the proposed system.

The OCM also has an Ethernet interface in order to manage the onboard subnet. On this subnet other devices can be connected, such as IP cameras, sensors, embedded systems or even PCs or laptops.

Using a host-to-host type Virtual Private Network (VPN) will cover two of the previously presented system contributions. On one hand, the communication will be encrypted with a cryptographic symmetric key, providing an additional security layer for data transmission between onboard and ground modules. Furthermore, the use of a VPN involves the creation and use of a virtual interface whose IP address will be the same regardless of the physical link being used at any time. This means that the use of these virtual interfaces for the communication between the two extremes ensures that applications do not have to change their settings when the active physical channel is changed. This supposes a new abstraction layer for the applications working with this vehicular communications system. Therefore, each of the available mobile nodes will be identified always with the same IP address.

Since most of network traffic will be sent from the mobile nodes to the ground centre, a QoS management middleware will be implemented in the mobile end, setting the network traffic rules on the virtual interface created by the VPN. Thus, it does not matter what the currently active link is, since all network traffic will be transmitted using the virtual interfaces configured on the system.

As the available network bandwidth is not stable as it could be in a static network, the QoS becomes more important. The network consumption priorities should be managed and adapted each time the available bandwidth fluctuates.

Summing, the onboard module must behave like a kind of router able to: manage the host-to-host type VPN to ground module, manage the private subnet of the mobile node, prioritize outgoing network traffic, run third party software and redirect the data traffic of the private subnet to the ground centre.

A. Software utilities used

No software was developed in this proposed communications system, but it has tried to combine and configure already existing and available software tools to meet the contributions presented in Section 2.

TABLE I. MATCHING OF SYSTEM CONTRIBUTIONS AND THE SOFTWARE TOOLS USED TO THEIR FULFILMENT

Contribution	Technical solution and software utility used
Network traffic regulation (prioritization)	QoS management (iptables + Traffic Control)
Network security	Point-to-point VPN (OpenVPN)
Onboard subnet management	Network gateway (Webmin)

For the deployment of the system, the software used was (Table 1): Ubuntu 11.10 [19] as GNU/Linux distribution, OpenVPN [20] for the host-to-host VPN management and various utilities from the Iproute2 [21] utility collection and Netfilter framework [22], mainly iptables and Traffic Control for the QoS management.

1) Operating System (GNU/Linux)

Although it can be found equivalent tools on different operating systems like Microsoft Windows, it was decided to choose a GNU/Linux distribution for two reasons: first, that is free and open source, and second, that is easier to find and modify network management tools than in others.

2) VPN (OpenVPN)

As VPN management software, OpenVPN was used due to its ease installation and free use.

A host-to-host type VPN must be configured for each onboard module to be managed from the ground node. The latter is the responsible for managing the communications between the different mobile nodes if they wanted to make a communication from a mobile node to another.

This type of VPN requires that one of the two nodes acts as a server and the other one as a client. Considering that the onboard physical links will have variable IP addresses depending on which the current active link is and the location of the mobile node, the ground module will be the VPN server and will be in charge of receiving request for connection establishment from each of the physical interfaces installed in the onboard modules.

Therefore, the design of the network architecture follows a star topology, where the ground module is the central node of the graph and the mobile modules are the leaf nodes.

The client-server connection establishment is made using the default route defined by the routing table of operating system. This can be modified using the ip route command, from the Iproute2 utility collection [21], available in most of the GNU/Linux distributions. In case of modification of the default route, OpenVPN detects it and manages the reconnection to the server using the new route.

3) Quality of service (iptables and Traffic Control)

To ensure the quality of service of the active channel, a combination of iptables and Traffic Control utilities has been used (Figure 2).

Iptables belongs to the framework Netfilter and it is the default firewall used in GNU/Linux. For this solution, its packet marking module will be used. With this module a mark will be added to each data packet redirected to the external network, either from the private subnet or the system itself. This classification is based on the port used to transmit each of the packages, so the data packets of each application running in the mobile node can be classified.

The data packet marking rules are easily configurable and replaceable in case of making changes on the system.

Traffic Control, from the utility collection Iproute2, will manage the queue disciplines, prioritizing the outgoing data traffic. After iptables has marked the data packets according to established rules, Traffic Control associates each mark to a priority class and then classifies and manages the bandwidth usage limits.

 ISBN: 978-1-61208-305-6

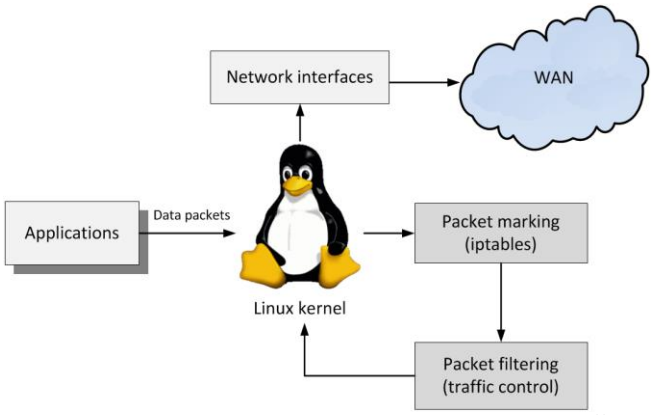


Figure 2. QoS management in GNU/Linux using iptables and Traffic Control utilities.

4) Queue disciplines (qdisc)

The queue disciplines determine the way in which data packets are sent. It is important to highlight that it can only be shaped the transmitted data and not the received one. Thus, the OCM will manage the queue disciplines to guarantee the QoS demanded by the applications and the adequate network usage prioritization.

Among the available queue discipline algorithms two have been considered for the proposed system: PRIO and HTB.

The PRIO qdisc (Priority queue discipline) does not need to define the current available bandwidth and it subdivides traffic based on how the Traffic Control filters are configured. It is a strong queue discipline for static networks in which the bandwidth fluctuates, such as neighbor shared network where the interactive traffic and non-interactive traffic must be managed, giving priority to the interactive one.

The Hierarchical Token Bucket (HTB) [23] queue discipline allows dividing the available network bandwidth indicating a maximum and a minimum usage for each application, ensuring that the highest priority applications of the system may have the required bandwidth at any time.

Despite of having to specify the available bandwidth each time the bandwidth changes, the HTB is proposed to be the queue discipline to use. We consider that it is a more adequate queue discipline in mobile networks mainly because it allows to configure the transfer rate per application. Moreover, it can be easily reconfigured with a few commands when the bandwidth changes; therefore, it does a better network prioritization than the PRIO queue discipline.

5) Gateway configuration (Webmin)

Ubuntu, as all other GNU/Linux distributions, can be configured to act as a network gateway. However, in order to facilitate this task and to provide a more user-friendly gateway, it was decided to use a web-based interface for system administration. To do this, it was chosen Webmin [24], which has all the necessary features, such as DNS and DHCP server.

The gateway was configured to forward the traffic from the subnet to the VPN tunnel and to apply the previous specified QoS rules in order to shape the network traffic.

V. TESTS SET-UP

In order to test our communication system, we have developed a simple application which triggers petitions from an onboard device to the GCM (Figure 3). In this communication, the traffic is forwarded to our system and it is shaped and prioritized according to its predefined configuration. This test application was run in a laptop which was connected to the onboard Ethernet network, so we could test two parts of the system: the network management and network prioritizing method.

It is important to highlight that this simple application is the unique software developed in this project and it has been used only to perform the tests. All the software that composes the solution already existed and was developed by third parties.

Moreover, only for informative purposes, during the tests the geolocation of the vehicle was captured using a standalone GPS device.

To verify that the proposed system works, a test plan has been developed and performed in laboratory settings, using a PC and an embedded system to simulate ground centre and an onboard module. Both systems have Ubuntu 11.10 and OpenVPN installed.

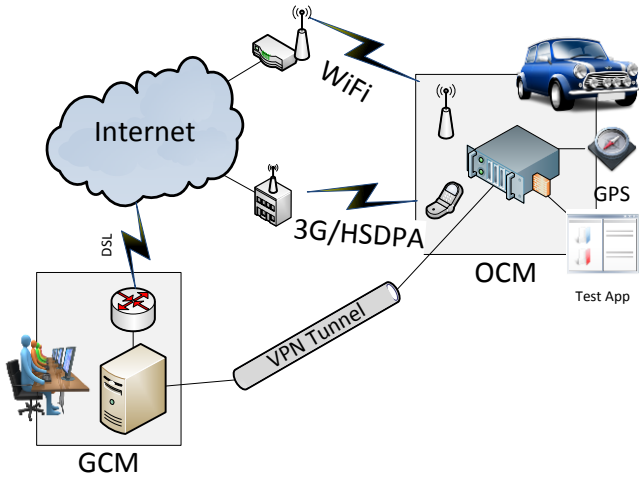


Figure 3. Conceptual diagram of the test implementation.

Summarizing, in this tests some files will be transmitted from the vehicle to ground, each one with different priority. The goal is to validate that the system performs properly. It means that all the requests are transmitted in compliance with the established minimum transfer rate and in the case of having free bandwidth it will be assigned to the highest priority request.

A. Scenario configuration

The system was tested in a real vehicular scenario. Although a system like the proposed one actually would be deployed into a public transportation vehicle, the tests were

 ISBN: 978-1-61208-305-6

designed to be performed in a private car seeking to emulate the same conditions as would occur in a public bus.

In this section, a description of the geographical scenario in which the tests has been performed is described. There are three main elements in the scenario configuration: the path, the vehicle, and the network link.

In the selection of the scenario path, it was considered to have a bandwidth fluctuating scenario, so a mixed urban and outskirts path was chosen. The path goes from the University of Deusto (Bilbao, Spain) to the beach of Sopelana (Spain), located to 16km away (Figure 4).

As a vehicle, it was used a common private car, a Renault Clio from 2008, and the travel was made in an average velocity of 80Km/h, as it would be in a public bus. In addition, two people were required in the car, one driving and the other one supervising the embedded system and the test application running in a laptop.

Due to the chosen queue discipline behaviour, the system must know the available bandwidth in each moment so that the traffic prioritization is done as intended. For this purpose, before doing the real vehicle travel and test, a current-available-bandwidth-capture was done. Thus, it was used three 3G USB dongle from different Internet Service Providers. The bandwidth data was captured using Iperf [25], a free and open source network tool. Iperf can also provide more data of the network, such as network latency, but for these tests we only needed to use this tool to get the available bandwidth on each moment.

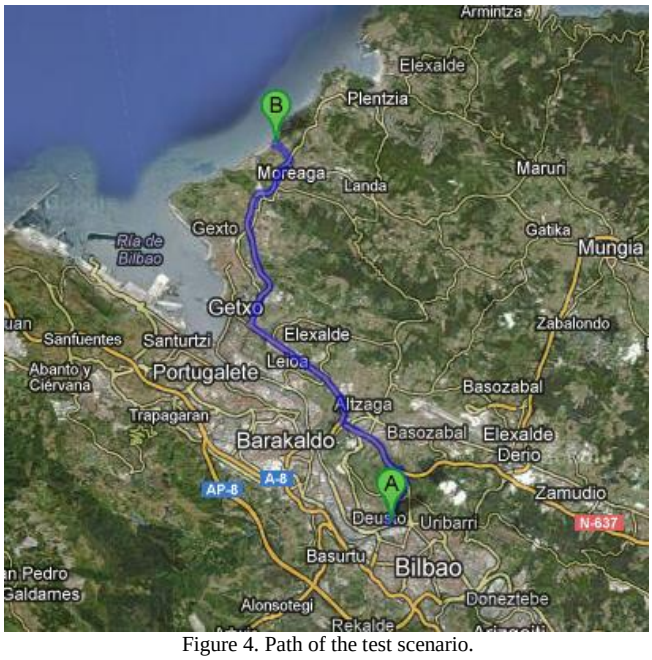


Figure 4. Path of the test scenario.

B. *Running the tests*

Once the first bandwidth-catching trip was done, the link with more bandwidth changes was chosen. So, the proposed system would be tested in the worse possible scenario to get the best analysis about the proper performing of the communication system.

Having in mind that the system has to know how much bandwidth is available in each moment and assuming that the available bandwidth would be similar to the data captured previously, some scripts were prepared by which the system changed the network prioritization adapting its configuration to the current network status.

Each test was composed of four requests. Each request with a different level of precedence: low, normal, high and priority. The planning of requests (minutes when they are triggered) was the following:

- Minute 0: normal priority request.
- Minute 1: high priority request.
- Minute 2: low priority request.
- Minute 3: the highest priority request.

This test suite was done repeatedly along the path to the end of the trip, so results of different areas can be analysed after the tests execution.

VI. TEST RESULTS

To be able to analyse the test results, the developed testing application, every three seconds, logged the transfer rate of each request and a GPS device captured the position of the vehicle. Thus, we could identify the behaviour of the proposed system at any time and location.

Taking the graph showed in Figure 5 as our first test result set, we see that a total bandwidth of 120KB/s is assigned. In this situation, the bandwidth is divided according to the following priority levels and the minimum transfer rates needed:

- The highest PRIORITY request: 70KB/s
- HIGH priority request: 30KB/s
- NORMAL priority request: 15KB/s
- LOW priority request: 5KB/s

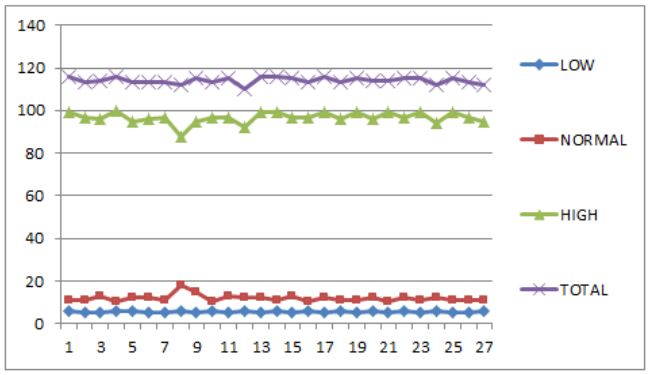


Figure 5. Data transfer division with three requests in a test chunk.

In this chunk of 27 seconds, it is shown that there are three requests running at this moment: a low priority request, a normal priority request and a high priority request. The highest priority request has previously finished so there is 70KB/s free bandwidth available. This free bandwidth is assigned to the high priority request for being the next in the priority list, and the other requests continue with the assigned transfer rate limit. Thus, it can be seen that the request with high priority has a 100KB/s transfer rate.

 ISBN: 978-1-61208-305-6

The results indicate that in most of the cases the prioritization of the network works as intended: the configured minimum transfer rate is complied; ensuring that every request meet the quality of service requirements and when there is free bandwidth it is divided according to priorities level.

Anyway, there are some cases in which the bandwidth is not divided as it has to and it is divided in an equitable way.

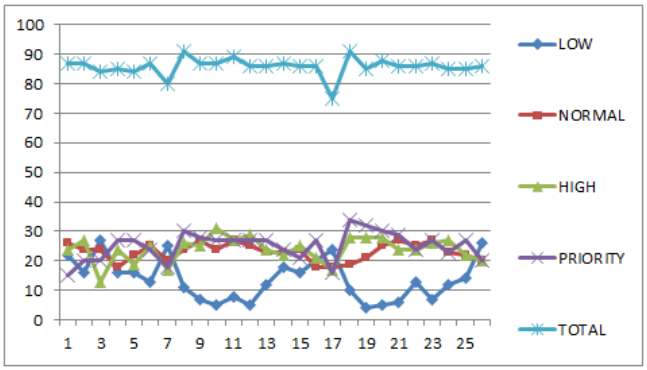


Figure 6. Data transfer division with four requests in a test chunk.

In Figure 6, we can see that the system has 90KB/s of total bandwidth assigned. With four requests of different priorities running in the test scenario, the bandwidth is not divided in a percentage, as it happened in Figure 5, and it is divided in a equitable way. Moreover, it can be seen that the LOW priority request sometimes is limited to its transfer rate, but it is not as constant as it has to be.

This equitable bandwidth division occurs when there is less bandwidth available than the specified one, so the queue disciplines cannot work as they are designed and solve the situation distributing the transfer rate in this way.

According to an intensive analysis of the previously presented test results and the system performance, we can confirm that the proposed communications system has some limitations that must be taken into account in case of a real industrial deployment. For this purpose, the system should be extended adding the abilities described in this section.

Due to the requirements of the chosen queue discipline in the network prioritization, it is necessary to know the bandwidth available at any time. This can be achieved by the method used in this paper, knowing in advance the bandwidth available in each section of the travel path. But this approach only has sense in tests scenarios or in very predictable ones. A more realistic solution could be to have a network monitoring tool that calculates the available bandwidth and updates the network configuration when the bandwidth fluctuates.

It should be noted that this kind of QoS systems (those supported by the set of Linux based software utilities used in this work) were designed to work in static environments, so the development and usage of this monitoring tool is absolutely necessary to adapt it to the current mobile environment. This limitation can be seen in the second result graph (Figure 6) in which the real bandwidth is lower than the specified one, so the queue discipline does not work as it is needed in this vehicular system. GNU/Linux does not have a dynamic QoS system developed [5], so having a network monitoring tool could be a solution to achieve it.

There is another improvement to be considered in this system: the continuous communication. Unlike the common network configurations used in non-mobile environments, such as in an office (where static network links are used, continuous communication can be assured using wired and backup links [2], and the service failure probability depends on rare environmental factors and internet service provider quality), in mobile environments the continuous communication is not as easy to guarantee. The reason is that the communication could be affected by several previously explained dynamic factors.

The best way to assure the network availability is having several 3G modems connected to the system, so if the active link is cut the system can choose another link to continue the communication. In order to accomplish this improvement, the already implemented VPN can be used. The host-to-host VPN tunnel provides an additional abstract layer to the onboard running applications, so after the active link changes the applications will run using the same IP as before the communication link has changed. Furthermore, due to this abstract layer, the link changes would not be detected as a broken link by the onboard applications, thus, the continuous communication between vehicle and ground should be achieved.

VII. Conclusions and Future Work

In this paper, we have presented the results of two years of research into the design and evaluation of a software system to provide a QoS management solution in dynamic network environments. The tests set-up in real vehicular scenarios, the execution process, the obtained results, and the consequent analysis has been also presented.

The communications system has resulted to be effective in the way in which prioritizes the network traffic, but it has some flaws that have to be fixed to be a fully operational system (in real deployments). Moreover, the lack of ability to ensure continuous communication and to provide an effective active channel change management should be fixed to be a complete vehicular network management system. This evolution of the system should be able to control the QoS of the onboard network and assure continuous communication between vehicles and the traffic control.

Using GNU/Linux utilities, an approximation to the continuous communication challenge could be achieved using the abstraction layer provided by the VPN tunnel, but another network monitoring tool would be needed to change the active link whenever the network coverage is lost.

Related work exists in the area of continuous communication in vehicular environments and also in network request prioritization. This is the case of the software tool [26] developed by our research group, which works as a middleware for monitoring the bandwidth consumption and available mobile network links and subsequently, managing the active channel change.

Nevertheless, the work presented in this paper is also focused on QoS management and prioritization of communications, but reusing already third party developed

 ISBN: 978-1-61208-305-6

software and using GNU/Linux as operative system. In conclusion, the objective was to obtain similar results from a new perspective, with GNU/Linux and the tools developed by the open source community in order to reduce programming efforts.

The future work will be focused on two areas. First, on the development of a tool able to dynamically monitor the bandwidth of the network links, adapting the QoS management rules to the bandwidth available in each moment. Second, adding continuous communication abilities to the system, by the adaptation and integration of one of our already existing research projects in this GNU/Linux based system.

Finally, there is also pending work relative to testing the active link change in real scenarios using a VPN as an abstract network layer [27]. It would be the best solution for a GNU/Linux based QoS and network management system as the one proposed here.

ACKNOWLEDGMENT

This work has been funded by the Basque Government of Spain under GAITEK funding program (GEINFEVI project, IG-2011/00472). Special thanks to DATIK - Irizar Group for their support.

REFERENCES

[1] D. Staessens, D. Colle, M. Pickavet, and P. Demeester, "Computation of high availability connections in multidomain IP-over-WDM networks", ICUMT '09, International Conference on Ultra Modern Telecommunications & Workshops, Oct. 2009, pp. 1-6.

[2] I. Salaberria, U. Gutiérrez, R. Carballedo, and A. Perallos, "Wireless Communications Architecture for "Train-to-Earth" Communication in the Field of Railways", DCAI, 2nd International Symposium on Distributed Computing and Artificial Intelligence, Jan. 2009, pp. 625-632.

[3] J. K. -S. Lau, C. -K. Tham, and T. Luo, "Participatory Cyber Physical System in Public Transport Application", UCC, Fourth IEEE International Conference on Utility and Cloud Computing, Dec. 2011, pp. 355-360.

[4] U. Gutiérrez, I. Salaberria, A. Perallos, and R. Carballedo, "Towards a Broadband Communications Manager to regulate train-to-earth communications", MELECON, 15th IEEE Mediterranean Electrotechnical Conference, Apr. 2010, pp. 1600-1605.

[5] X. Liu, "Supporting dynamic QoS in Linux", RTAS, 10th IEEE Real-Time and Embedded Technology and Applications Symposium, May 2004, pp. 246-254.

[6] L. Qi, "Research on Intelligent Transportation System Technologies and Applications", Workshop on Power Electronics and Intelligent Transportation System, Aug. 2008, pp. 529-531.

[7] F. Benzi, G. S. Buja, and M. Felser, "Communication architectures for electrical drives", IEEE Transactions on Industrial Informatics, Feb. 2005, vol. 1, pp. 47-53.

[8] M. Felser, "Real-time ethernet - Industry prospective", Proceedings of the IEEE, 2005, vol. 93, pp. 1118-1129.

[9] R. Ernst, G. Spiegelberg, T. Weber, and H. Kopetz, A. Sangiovanni-Vincentelli, and M. Jersak, "Automotive networks: Are new busses and gateways the answer or just another challenge?". CODES+ISSS: International Conference on Hardware/Software Codesign and System Synthesis, Salzburg, Sept. 2007, pp. 263.

[10] S. Kurowski, J. Zibuschka, H. Roßnagel, and W. Engelbach, "A Concept for Interoperability of Security Systems in Public Transport", Proc. of the 9th International ISCRAM Conference, Apr. 2012.

[11] P. Dharwadkar, H. J. Siegel, and E. K. P. Chiong, "A Heuristic for Dynamic Bandwidth Allocation with Preemption and Degradation for Prioritized Requests", ICDCS, 21st International Conference on Distributed Computing Systems, Apr. 2001, pp. 547-556.

[12] P. Jayachandran and T. Abdelzaher, "Bandwidth Allocation for Elastic Real-Time Flows in Multihop Wireless Networks Based on Network Utility Maximization", 28th International Conference on Distributed Computing Systems, June 2008, pp. 849-857.

[13] D. Marrero, E. M. Macias, and A. Suarez, "Dynamic Traffic Regulation for WiFi Networks", Proc. of the World Congress on Engineering, July 2007, pp. 1512-1517.

[14] M.F. Horng, Y.H Kuo, L.C. Huang, and Y.T. Chien, "An Effective Approach to Adaptive Bandwidth Allocation with QoS Enhanced on Ip Networks", ICUIMC, International Conference on Ubiquitous Information Management and Communication, 2009, pp. 260-264.

[15] P. Noh-sam and L. Gil-Haeng, "A framework for policy-based sla management over wireless LAN", 2005, Proc. of the Second International Conference on e-Business and Telecommunication Networks, INSTICC Press, ISBN 972-8865-32-5, pp. 173-176.

[16] I. Martínez, "Contribuciones a Modelos de Tráfico y Control de QoS en los Nuevos Servicios Sanitarios Basados en Telemedicina", Ph.D Thesis, Universidad de Zaragoza, 2006.

[17] I. Salaberria, R. Carballedo, and A. Perallos, "Wireless Technologies in the Railway: Train-to-Earth Wireless Communications", Wireless Communications and Networks - Recent Advances, Ali Eksim (Ed.), DOI: 10.5772/35962, ISBN 978-953-51-0189-5, March 2012, pp. 469-492.

[18] Boss: On Board Wireless Secured Video Surveillance http://celtic-boss.mik.bme.hu/ [retrieved : October 2013]

[19] Ubuntu GNU/Linux: http://www.ubuntu.com/ [retrieved : July, 2013]

[20] OpenVPN, VPN management software: http://www.openvpn.net/ [retrieved : July, 2013]

[21] Ip route, from Iproute2 collection utility: http://www.linuxfoundation.org/ [retrieved : July, 2013]

[22] Netfilter, packet filtering framework: http://www.netfilter.org/

[23] J. L. Valenzuela, A. Monleon, and I. San Esteban, "A hierarchical token bucket algorithm to enhance QoS in IEEE 802.11: proposal, implementation and evaluation", VTC, IEEE 60th Vehicular Technology Conference, Sept. 2004, vol. 4, pp. 2659-1662.

[24] Webmin: http://www.webmin.com/ [retrieved : July, 2013]

[25] S. S. Kolahi, S. Narayan, D. D. T. Nguyen, and Y. Sunarto, "Performance Monitoring of Various Network Traffic Generators", UKSim, 13th International Conference on Computer Modelling and Simulation, Apr. 2011, pp. 501-506.

[26] I. Salaberria, A. Perallos, and R. Carballedo, "Towards a Dynamic and Adaptative Prioritization of Wireless Broadband Vehicle-to-Ground Communications", ACCESS, The 3th International Conference on Access Networks, June 2012, pp. 31-34.

[27] G. Urquiola, A. Perallos, and R. Carballedo, "Continuous Broadband Communication System Base on Existing Open Source Network Tools for Vehicular Environments", ITSC, 15th International IEEE Conference on Intelligent Transportation Systems, Sept. 2012, pp. 248-253.

 ISBN: 978-1-61208-305-6

Analysis of PLC Channels in Aircraft Environment and Optimization of some OFDM Parameters

Thomas Larhzaoui, Fabienne Nouvel, Jean-Yves Baudais
IETR
Rennes, France
thomas.larhzaoui@insa-rennes.fr, fabienne.nouvel@insa-rennes.fr, jean-yves.baudais@insa-rennes.fr

Virginie Degardin, Pierre Laly
IEMN
Lille, France
virginie.degardin@univ-lille1.fr, pierre.laly@univ-lille1.fr

Abstract— **PLC technology based on an OFDM communication scheme is considered for data transmission between a control unit and actuators located in an aircraft wing. In order to optimize some OFDM parameters, transfer function measurements have been performed on an avionic test bench. Based on experimental values of the coherence bandwidth and of the channel delay spread, it appears that the subcarrier spacing can be larger than the spacing specified in HomePlug AV, which usually applies for in-house environment. Similarly, the duration of the cyclic prefix can be reduced. This will allow meeting the real-time and determinism constraints of avionic systems.**

Keywords-PLC, OFDM, coherence bandwidth, dela spread, insertion gain, aircraft

I. INTRODUCTION

In future aircrafts, hydraulic flight control systems will be replaced by electric systems. The main interests are: a better reliability and flexibility, a decrease in maintenance costs, but the major problem is the increasing of wires length. In order to decrease this length, it has been proposed to use power line communications (PLC) technology for flight control systems. This technology has proven its reliability in in-home network with HomePlug AV (200 Mbit/s in the [1;30] MHz bandwidth [1]). In addition, there are numerous studies concerning PLC in different kinds of vehicles like cars [2], boats [3], and trains [4]. For aircrafts, [5] and [6] projects investigate the possibility of using PLC technology for the cabin light system.

In the near future, the AC power distribution will be replaced by a high voltage direct current (HVDC) distribution network. The authors in [7] studied the feasibility of using PLC technology between the power inverter and the actuator for flight control. In this paper, we focus on the PLC link on the HVDC network between a control unit and the power inverter feeding various active loads. Flight control systems do not require a high bit rate link, few Mbit/s being enough, but the communication must be highly reliable, deterministic, real time and must comply with the aeronautic standard DO-160 [8]. As shown in Fig. 1, we will consider the link between the control unit and the power inverter located near the actuators (spoilers) used for flight control. In this illustration, the PLC master (control unit side) transmits data to PLC slaves (power inverters side), corresponding to a point-to multipoint architecture. It is also possible to use point-to-point architecture, where one PLC node transmits data to another PLC node.

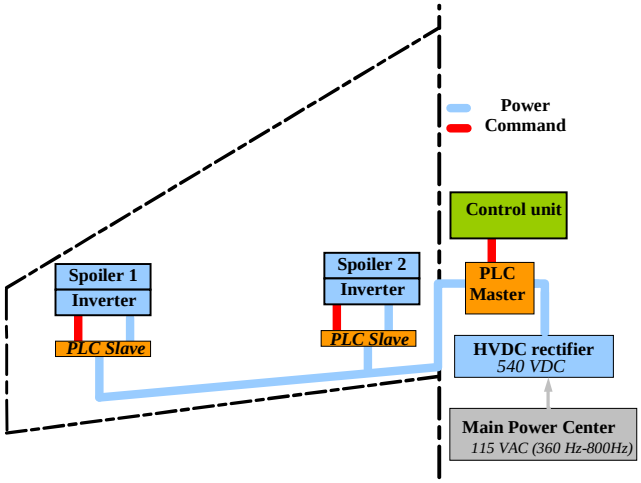


Figure 1. PLC system on aircraft wing

In order to transmit data over the HVDC network, we use either a capacitive coupler or an inductive coupler, the transmission being based on the usual Orthogonal Frequency Division Multiplexing (OFDM) transmission scheme. In addition, one of the major challenges of the command of the flight control systems is the real time constraints. Indeed, they operate at frequency about 1kHz (fast loop). According to the common practice, command systems must work six times quickly as much as the equipments that they command, which represent 6 kHz in our case. In addition, we must pay attention that there are several calculators in the loop, which require time processing. That's why we consider that our system do not exceed from 10 % to 20 % of the total time processing, which represent in our case from 17 µs to 34 µs. Thus, an analysis of the propagation channel made on a test bench representative of an actual configuration will allow the optimization of few OFDM parameters to show that is possible to have an OFDM duration in compliance with these real-time constraints.

This paper is organized as follows. In Section II, we describe the channel and the test bench, while results on the

 ISBN: 978-1-61208-305-6

insertion gain are presented in Section III. Section IV, describes the channel analysis and the optimization of the OFDM parameters is presented in Section V. A synthesis of the main results and a conclusion are given in Section VI.

II. DESCRIPTION OF THE TEST BENCH AND OF THE MEASUREMENT PROCEDURE

In this test bench, the channel is formed by a harness and at least 2 couplers.

A. *Configuration*

In this measurement campaign, two architectures were studied:

- The point-to-point architecture, with 2 couplers (Fig. 2 and Fig. 3)
- The point-to-multipoint architecture with 1 master and 2 slaves (Fig. 4)

The tests have been performed on a test bench whose active loads (540 V, 5 A) are representative of actual avionic loads and with a ± 270 DC power supply.

For this experimentation, the harness of 32 meters long, includes one twisted pair, one twisted quadrifilar cable, and one wire. The capacitive coupler transmits data between +270 V and -270 V DC, and is made by a transformer and two capacitors and use one twisted pair: signal is transmitted on one twisted pair between +270 V and -270 V.

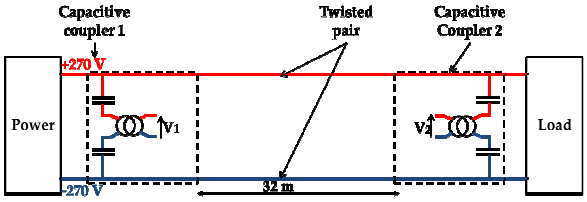


Figure 2. Point-to-point architecture with capacitive coupler

We have also considered another possibility, which is to use a twisted pair for the +270 V, rather than a single wire for the -270 V. In this case, the twisted pair is short circuited at both ends and an inductive coupler can be used as illustrated in Fig. 3: signal is transmitted on one twisted pair on +270 V. It must be emphasized that, for the same DC power, the diameter of each wire of the twisted pair can be reduced for avoiding an increase of the weight.

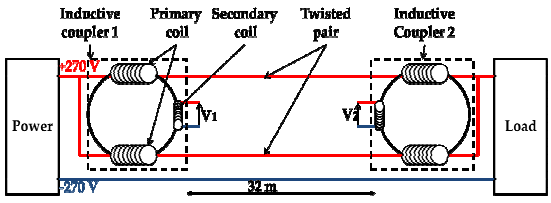


Figure 3. Point-to-point architecture with inductive coupler

In Fig. 4 we shows the last possibility: the point-to-multipoint architecture with inductive coupler. In this case we use three couplers on the +270 V. thus, harness is compose of one quadrifilar and one wire for the minus polarity. In the following, line 1 refers to the channel between coupler 1 and coupler 2 and line 2 to the channel between coupler 1 and coupler 3.

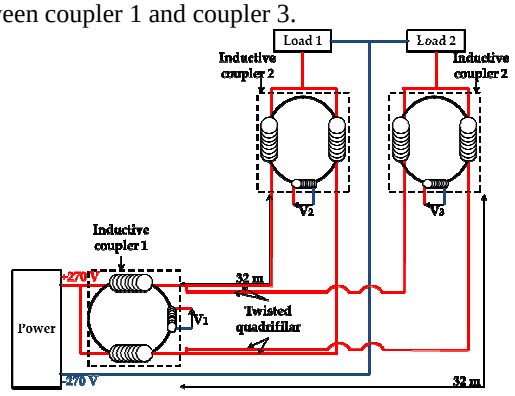


Figure 4. Point-to-multipoint architecture

B. *Measurements*

Measurements have been carried out with a network analyser in the [1;100] MHz bandwidth with a 5 kHz resolution bandwidth. For each configuration, the transfer functions were measured between the input/output V_1, V_2 and V_3.

III. INSERTION GAIN AND CHANNEL IMPULSE RESPONSE

Plots of Fig. 5 give the variation of the insertion gain (IG) versus frequency for the different architectures, while the cumulative distribution of IG is represented in Fig. 6. For the point-to-point configuration with an inductive coupler, IG first decreases linearly (in dB) with the frequency (up to 40 MHz), and varies from -5 to -25 dB. Then, IG remains nearly constant between 40 and 80 MHz and, beyond 80 MHz, decreases very rapidly. The other plots of Fig. 5 show that this behavior is nearly independent from the configuration except that we can note that the point-to-point link with a capacitive coupling presents several important fading in the low frequency band while at high frequency, (between 70 and 100 MHz), IG takes higher values than for the other configurations.

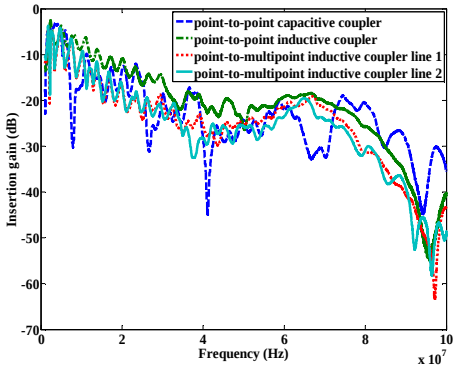


Figure 5. Insertion gains for the different configurations

In Fig. 6, probability at 10% is about -40 dB for inductive coupler and about -30 dB for capacitive coupler. The

probability at 50 % is about -20 dB for point-to-point architecture with inductive coupler and about -25 dB for the others configurations.

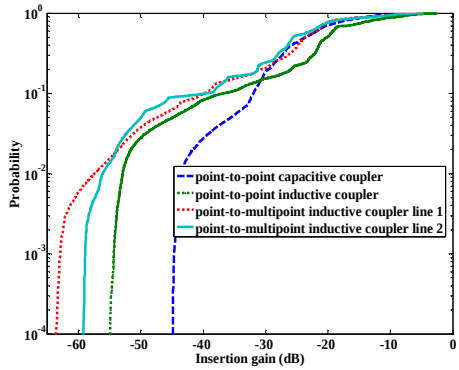


Figure 6. Cumulative distribution function of insertion gain

The channel impulse response has been deduced from the measurements of the complex transfer function by applying an inverse Fourier transform of 6000 points for [1;30] MHz and 20000 points for [1;100] MHz bandwidth. The results are shown in Fig. 7 and Fig. 8, by considering a frequency band either between 1 and 30 MHz or between 1 and 100 MHz.

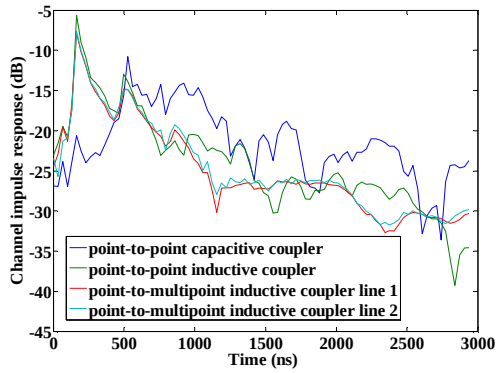


Figure 7. Channel impulse response in the [1;30] MHz

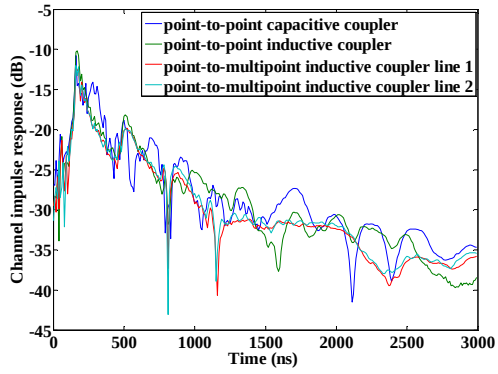


Figure 8. Channel impulse response in the [1-100] MHz

Coherence bandwidths and delay spreads are deduced from these results in frequency and time domain.

IV. COHERENCE BANDWIDTH AND DELAY SPREAD

The coherence bandwidth (CB) is deduced from the absolute value of the autocorrelation of the complex transfer function [9]. In the following, CB is calculated for a correlation coefficient of 0.9. Delay spread is calculated from the channel impulse responses according to [9]. Delay spread and coherence bandwidth are given in Table I, by considering either a 30 MHz or a 100 MHz transmission bandwidth.

It appears that the geometrical architecture and the type of coupler do not have a strong impact, the coherence bandwidth being of the order 700 – 1000 kHz, while the delay spread varies from 78 to 104 ns. These results are quite comparable to those obtained for other embedded systems as shown in Table II.

TABLE I. COHERENCE BANDWIDTH AND DELAY SPREAD FOR DIFFERENT CONFIGURATIONS

	Delay spread (ns)		**coherence bandwidth 0.9 (MHz)**	
Configurations	***[1;30] MHz***	***[1;100] MHz***	***[1;30] MHz***	***[1;100] MHz***
Capacitive coupler in point-to-point mode	*104*	*79*	*0,70*	*0,72*
Inductive coupler in point-to-point mode	*83*	*78*	*0,91*	*1,02*
Inductive coupler in point-to-multipoint mode (line 1)	*99*	*103*	*0,72*	*0,80*
Inductive coupler in point-to-multipoint mode (line 2)	*97*	*87*	*0,71*	*0,77*

TABLE II. COHERENCE BANDWIDTH AND DELAY SPREAD FOR DIFFERENT VEHICLES

References	**Bandwidth (MHz)**	**Delay spread (ns)**	**Coherence bandwidth (MHz)**
[2] car	[1-50]	[34-200]	[0.4-4.8]
[7] aircraft	[1-30]	100	[0,6-0,9]
[10] car	[0.3-100]	130	0.48
[11] car	[1-70]	380	[0.4-0.7]

V. OPTIMIZATION OF FEW OFDM PARAMETERS

A. *OFDM Subcarrier Spacing*

In order to meet the real time constraints, it is necessary to minimize the processing time of the data. Since fast Fourrier transform (FFT) is a time consuming process, one can try to decrease the number of carriers and choose, as in common practice, a carrier spacing equal to about 10% of the coherence bandwidth. Taking the values in Table I into

 ISBN: 978-1-61208-305-6

account, this leads to a 70 kHz subcarrier spacing, thus about three times the value given in HomePlug AV specifications (24.414 kHz). Furthermore, the baseband complex OFDM symbol will supply an I/Q RF modulator. This allows us to keep the FFT size equal to the number of carriers. In our case, the number of subcarriers is equal to 428 or 1428 for a transmission bandwidth of 30 MHz or 100 MHz, respectively.

B. *Interference Characterization*

Using the channel impulse response values, it is also possible to compute the Inter Symbol Interference (ISI) and the Inter Carrier Interference (ICI) in order to choose the optimal length L_{cp} of the cyclic prefix. A too long L_{cp} will reduce spectral efficiency and data rate. The power spectral density of (ISI) and (ICI) can be computed as follows [12]:

$$N_{ISI+ICI}(n) = 2\sigma_x^2 \sum_{l=L_{cp}+1}^{L_c-1} \left| \sum_{u=l}^{L_c-1} h(u) \exp\left(-j \frac{2\pi}{N} un \right) \right|^2 \quad (1)$$

In (1), σ_x^2 is the variance of modulated signal (and normalized to 1 W), h is the channel impulse response, L_c the channel length expressed in number of samples, L_{cp} being also expressed in terms of number of samples, N the number of carriers, and n the subcarrier index.

Fig. 9 and Fig. 10 give the interferences level, expressed in dBm/Hz for the inductive coupler in point-to-point configuration, and calculated in the [1;30] MHz and in the [1;100] MHz bandwidth, respectively. $N_{ISI+ICI}$ has been plotted versus the subcarrier number and for various lengths of the cyclic prefix (CP).

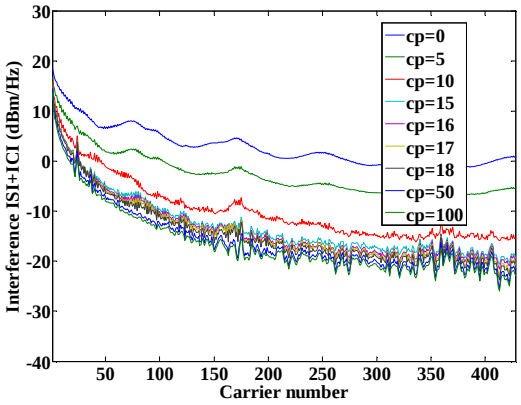


Figure 9. Interference in the [1-30] MHz bandwidth

As expected, the interference decreases rapidly with the length of the cyclic prefix but, beyond a given value, it does not vary appreciably.

From these curves, which have also been plotted for the other previously described network architectures, one can conclude that the CP length L_{cp} can be reduce at 15 samples (500 ns) for a 30 MHz band, and of 30 samples (300 ns) for a 100 MHz band.

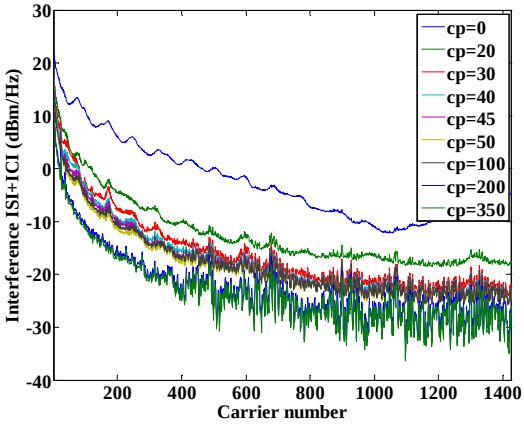


Figure 10. Interference in the [1-100] MHz bandwidth

As a comparison, if the cyclic duration was chosen equal to 2 to 4 times the delay spread, as suggested in [13], we would obtain a CP duration between 200 and 400 ns.

VI. SYNTHESIS AND CONCLUSION

For the aircraft environment studied in this paper, we have shown that it is possible to reduce the duration of an OFDM symbol compared to Homeplug Av standard, by increasing subcarrier spacing and decreasing the cyclic prefix duration. In the HomePlug AV standard the subcarrier spacing is 24.414 kHz, the minimum cyclic prefix duration is 5.56 µs, and the OFDM duration is 46.52 µs. In our application, we propose to increase the subcarrier spacing to 70 kHz and decrease the CP duration to 500 ns. Consequently, the symbol duration will be 14.78 µs instead of 46.52 µs (HPAV spec.). These results will help us to define the physical layer for a PLC avionics system in accordance with real-time constraints of a fast loop. This study can be applied to other critical avionic systems running on a HVDC network like landing gear. A fortiori, it is possible to use this study for a slow loop on HVDC network like, thrust reversal.

ACKNOWLEDGMENT

This paper is funded by Sagem and the harness was provided by Safran Engineering Services.

 ISBN: 978-1-61208-305-6

REFERENCES

[1] HomePlug Av specification, Version 1.1, May 21, 2007

[2] P. Tanguy and F. Nouvel, "In-vehicle PLC simulator based on channel measurements," International conference on intelligent transport system telecommunication (ITST),2010 , pp. 1-5.

[3] S. Barmada, L. Bellanti, M. Raugi, and M. Tucci, "Analysis of power-line communication, Channels in Ships," IEEE Trans. on Vehicular Technology, vol. 59, no. 7, september. 2010, pp. 3161-3170.

[4] S. Barmanda et al., "Design of a PLC system onboard trains: selection and analysis of the PLC channel," ISPLC, 2008, pp. 13-17.

[5] S. Bertuol et al, "Numerical Assessment of Propagation Channel Characteristics for Future Application of Power Line Communication in Aircraft," 10th Int. Symp. on EMC, September. 2011, pp. 506-511.

[6] V. Degardin, I. Junqua, M. Lienard, P. Degauque, and S. Bertuol, "Theoretical approach to the feasibility of power-line communication in aircrafts," IEEE Trans. VT, March. 2013, vol. 62, no. 3, pp. 1362-1366.

[7] K. Kilani, V. Degardin, P. Laly, M. Lienard, and P. Degauque, "Impulsive noise generated by a pulse width modulation inverter : modeling and impact on powerline communication," ISPLC, March. 2013, pp. 75-79.

[8] Do160,Environmental conditions and test procedures for airborne equipment, Standard, 2007.

[9] T.S. Rappaport, Wireless communication principle and practice, prentice hall patr,1996.

[10] A. B. Vallejo-Mora, J. J. Sanchez-Martínez, F. J. Cañete, C. J. Antonio, and L. Díez, "Characterization and evaluation of in-vehicle power line channels," in the IEEE Global Telecommunications Conference, 2010, pp. 1-5 , 2010.

[11] M. Lienard, M. Olivas Carrion, V. Degardin, and P. Degauque, "Modeling and analysis of in-vehicle power line communication Channels," IEEE transaction on vehicular technology, March. 2008 VOL. 57, NO. 2, pp. 670-679.

[12] W. Henkel, G. Taubock, P. Odling, P. Borjesson, and N. Petersson, "The cyclic prefix of ofdm/dmt - an analysis," in IEEE International Zurich Seminar on Broadband,Communications, 2002 pp. 22–1 –22–3.

[13] R. Van Nee and R. Prasad, "OFDM for wireless multimedia communications. Norwood," MA: Artech House, 2000.

 ISBN: 978-1-61208-305-6

Privacy-aware Nomadic Service For Personalized IPTV

Amira Bradai, Emad Abd-Elrahman, Hossam Afifi
RST. Telecom SudParis
Institut Mines-telecom, Telecom SudParis
Evry, France
amira.bradai@it-sudparis.eu, emad .abd-Elrahman @it-sudparis.eu, hossam. afifi@it-sudparis.eu

***Abstract*-User-Centric Personalized IPTV UbiquitOus and SecUre Services (Up-TO-US) project provides nomadism and personalization in IPTV operated services. This paper proposes a novel architecture for nomadism combined with an extensive game to enforce identity exposure when a user accesses his/her services from outside the home domain. The goal of the game is to minimize the personal information divulgation outside his/her domain. The proposed algorithm is implemented within a nomadic architecture. For each client, our algorithm can customize the Electronic Program Guide (EPG) in a contextualized way. We analyze two main use cases in nomadic situations: local nomadism and inter-domain nomadism. All implementations are hypertext Transfer Protocol (HTTP) standard-based for compatibility issues with all Internet Protocol Television (IPTV) platforms.**

Keywords-Nomadism; extensive game; IPTV; personalized service

I. INTRODUCTION

The advance in Internet Protocol Television (IPTV) technology enables a new model for service provisioning by moving from traditional broadcaster-centric TV services to a new user-centric TV model. This new model allows users not only to access new services and functionalities from their providers, based upon their profiles and contexts, but also to become active parts in the content personalization through contributing in building their dynamic profiles and their privacy. This IPTV model is promising in allowing low cost services for end-users and a revenue system for broadcasters based on personalized advertising methods, as well as new business opportunities for network operators and service providers.

A. UP-TO-US Project

The objective of the UP-TO-US project [1] is to elaborate, to prototype, and to evaluate an open European solution allowing IPTV services personalization over different IPTV systems (having different architectures and belonging to different network operators and service providers). The project assures content personalization according to each user context and the context of his environments (network and devices), while preserving his privacy, as shown in Figure 1.

UP-TO-US focuses on two use-cases for service personalization:

1. users in nomadic situations in a hotel, in a friend's home or anywhere outside his domestic sphere, (allowing the user to access his personalized IPTV content in a hotel for instance and be billed on his own bill "My Personal Content Moves with Me"), and
2. users' mobility in his domestic sphere (allowing the user to move around within his domestic sphere, while continuing accessing his IPTV service personalized according to his location and devices in his proximity "My Content Follows Me in a Customized Manner").

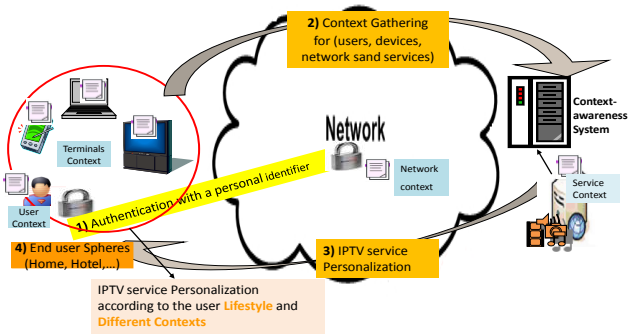


Figure 1. General visions of UP-TO-US project

In order to achieve the objectives of UP-TO-US, some enabler technologies were developed and integrated to different IPTV systems; these mainly include:

- A context-aware module capable of monitoring and gathering the user and his environment contexts and feed them in a dynamic manner to the IPTV system,
- A profiling management module, capable of constructing and dynamically updating the users profiles according to the various contexts, and

 ISBN: 978-1-61208-305-6

- A privacy management module that will be responsible for managing the different privacy levels for each user and protecting the user personal information. Consequently, personalized services could be provided in which content is selected according to users' preference, Quality of Experience (QoE) requirements, and different contexts, while fostering trust between viewers and broadcasters through an efficient privacy management, and thus encouraging viewers to participate actively in this interactive user-centric TV paradigm through allowing their continuous contexts gathering.

B. *Personalized EPG*

IPTV service personalization has been discussed in many works [7] [8] [9]. All those works focused on service personalization-based context-aware system.

With the new era of TV and different modes of diffusion like satellite/Triple and Quadruple play services offered by many service providers worldwide, there is a need to customize the huge number of channels. This customization is mandatory, especially in the nomadic places where users are away from their home networks.

Based on user profile, and on user and network gathered contexts, the personalization technologies can generate content recommendations for an individual as well as for a group of users. Recommendation algorithms can use content-based filtering technologies [12], collaborative filtering algorithms [13], or hybrid solutions combining content-based and collaborative filtering techniques [14][15].

To do this task, the recommendation system will be in charge of calculating the most interesting videos or channels for each client through her/his Electronic Program Guide (EPG). The EPG is the way to guide and inform the IPTV viewers by their interesting channels and programs. To customize an EPG in contextualized way is one of main objectives in UP-TO-US project. This customization is mainly depending on four types of context information, as shown in Figure 2:

- **User Context:** The information concerns the user location.
- **Network Context:** The information concerns the environment like network parameters.
- **Service Context:** The information concerns the service adaptation and its delivery status, coding, definitions (High Definition, Standard Definition or Low Definition).
- **Terminal Context:** The information concerns the device capabilities and its screen resolution.

A. *EPG format*

In UP-TO-US, we send and receive the EPG in XML format, as shown in Figure 3. The information in this list could be live TV channel, Video-on-Demand (VOD), programs or the Top-K favorites programs for the client.

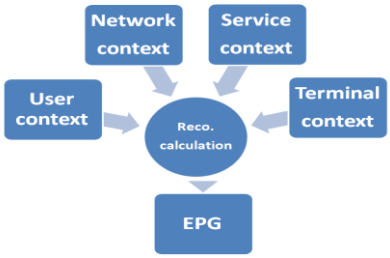


Figure 2. EPG calculation based context information feed to Recommendation System

```
<up2us:EPG>
<up2us:ListProgramDescription
xsi:schemaLocation="urn:tva:up2us:2012Up2us.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:tva="urn:tva:metadata:2011"
xmlns:mpeg7="urn:tva:mpeg7:2008"
xmlns:up2us="urn:tva:up2us:2012">
            <up2us:Options>Top_K</up2us:Options>
          <up2us:ProgramDescription>
                    <up2us:ProgramInformationTable
metadataOriginIDRef="IMDB">
                              <up2us:ProgramInformation
programId="crid://abc3/xyz24">
                                        <up2us:BasicDescription>
</up2us:EPG>
```

Figure 3. XML format for EPG in up-to-us

D. *Nomadic access*

Quality for Nomadic Access (NA) solutions are based on the network integration between different service providers (i.e., managed services operators) who are proposing key-point in geolocalisation solutions for video services delivery (IPTV or VOD) to their nomadic clients.

Customers are searching for service personalization anywhere/anytime/anydevice.

In a nomadic scenario, we may have two modes of service accessing, as shown in Figure 4:

- **Managed Mode:** the operators could guarantee the service for the clients according to the user service and context profile. In this scenario, the QoS/QoE trends can be adjusted and end-to-end controlled by network operators.
- **Unmanaged Mode:** which means accessing from Internet and in this case we must have Multi Criteria Decision Making (MCDM) for the quality and the cost to decide which is suitable for the user and which is suitable for the network. Based on the calculations and decisions conducted by this system (Decision Makers), the user can choose the best

 ISBN: 978-1-61208-305-6

service (which is a nomadic service) and the network could assign its best resources.

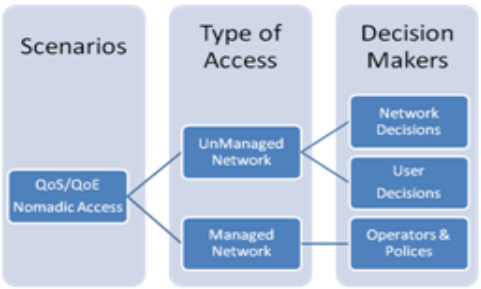


Figure 4. Quality aspect in IPTV nomadic access

The main objective of nomadic access for IPTV service is to overcome the challenges of obtaining the same service anywhere, anytime and on any personal device. Moreover, nomadic service adds the means of obtaining the service from any operator. The aspect of TV Everywhere [2] started some years ago and achieved high acceptance from many IPTV clients. Also, the report [2] showed that by soon, a large percent of the customers will able to access to TV Everywhere services through their current providers only.

The rest of this paper is organized as follows. Section II highlights the nomadic architecture components and the implemented scenarios in the project. Privacy aspects are handled in Section III. The proposed game solution is presented in Section IV with some highlighting on privacy in personalization. Then, the paper is concluded in Section V.

II. NOMADIC ARCHITECTURE

Nomadism is an equivalent term to roaming services, but without mobility [3]. Nomadic services allow the user to access to his personalized IPTV content in any place in his domestic and outside and to be billed on his own bill. The collaboration between operators helps in providing nomadic IPTV services in the context of cloud computing as explained by Abd-Elrahmen and Afifi in [4]. Moreover, the interconnection for IPTV terminals in visited networks with different services scenarios is explained in [5]. Through that, we can get the service either by home, visited or third party operator.

In UP-TO-US, the proposed components of the nomadic service and personalization are implemented in C language. We used HTTP to communicate with the context-aware and the user profile servers. The core network and IPTV service platform is achieved through HTTP and DIAMETER [6] protocols.

We achieved the nomadism through integrating Nomadic Service Module (NSM) in the IPTV platform. This NSM has two parts:

- **NSM client**: this part is integrated in the Set-Top-Box (STB) to add the nomadic features in the software (or hardware) client.
- **NSM server**: this server is the gateway function for nomadic services in operator network. Also, it is responsible for nomadic decisions, user domain searching and all inter-domain actions.

A. *Use Cases Analysis*

In our implementations, the client is software STB [1] developed by one of the project partners (Orange Labs). Then, we integrated our client module (NSM Client) in this STB. The client acts as hub for all requests and answers concern the nomadic access. The interconnection point for any type of access whatever nomadic or not is called IPTV Service Selection Function (SSF), which is also the gateway for IPTV platform.

In this part, we analyze two nomadic use cases; local and inter-domain as follows.

1) *Local Nomadism*

When the user resides outside his home but within his operator network, the NSM of visited and home network are the same. The initialization phase for searching user rights to access nomadic service in the visited locations is explained in the sequence shown in Figure 5. All messages sent and received in HTTP format to guarantee compatibility and standard issues. The client initiates the Nomadic Client Request (NCR) and waits the answer from the NSM server as Nomadic Service Answer (NSA). In case of faked login, the 'VALUE1' attribute will return 'NOT ALLOWED'.

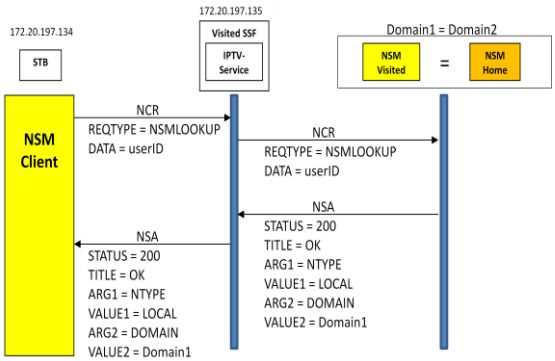


Figure 5. Users access their services outside their homes from their home operator (same domain)

2) *Inter-domain Nomadism*

Figure 6 illustrates the sequence diagram for initialization phase in case of inter-domain access. In this case, we have different domains; so, the visited NSM has the responsibility of searching which home network should be contacted.

 ISBN: 978-1-61208-305-6

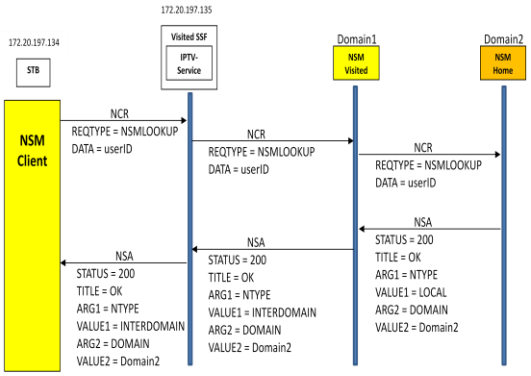


Figure 6. Users access their services outside their homes from another operator (different domain)

Then, the home NSM is responsible for service confirmation and checking users' rights with other modules in UP-TO-US.

III. PRIVACY ASPECTS

There exist many risks in storing personal information in online systems. These data may be used fraudulently to perform different attacks against its owners. Many devices may be used to deceive and defraud customers on social networks. There are many ways that information on social networks can be used for purposes other than what the user intended. Data can be obtained through many different means like for example eavesdropping.

Recommendation operations and context-aware system like the ones defined within our UPTOUS project rely on the use of an exhaustive amount of information provided implicitly or explicitly by the user. The more generous and detailed the personal data is, the more accurate the preferences inferred from the usage history are. In the case of our system, the information required from the user in order to perform a personalization operation covers a wide range of aspects. Many users choose to mask their real identities. Masking may be done via anonymity (providing no name at all) or pseudo-anonymity (providing a false name). By establishing anonymous profiles or alter egos, users decouple the different types of profiles they have defined. The main purpose is to keep a strict separation between the online person and the offline individual. However, it is still very difficult to separate the different identities or profiles defined for every user.

Privacy protection is a critical issue for personalized IPTV services acceptability by the users.

That is why the privacy layer in the proposed system supports a multi-identities mechanism. Each person holds multiple identities (each one with unique user Identity (Id)), depending on the user preferences about different data disclosure settings.

After authentication process has been accomplished for a subscriberId, a sessionId has been returned to the User Domain (UD).

Identities are used internally, while deciding which data is available for a personalization operation. UserId is only exchanged with the service domain in those interactions that involve users performing crud operations on identities. The Privacy Layer consists of:

- **Multi-identity Mechanism**: manages which identity (userId) is active at a moment for a person
- **Privacy Control/ Settings**: determines which information about the user behavior should be considered while inferring preferences
- **Users' Multi-Identities and Privacy Policies DB**: this database contains all data required by the privacy layer to perform its functionalities.

UP-TO-US has defined two different user domains: domestic and nomadic. Users in nomadic scenarios generally visit a platform different than the one they usually do; hence, they can't be sure about who else is using the same network and which information is available for others. Therefore, situations where nomadic scenarios are involved require a higher level of privacy than domestic situations. Privacy is one of the most important issues in our evolving information age, where technological developments lead to intensive processing and storage of personal information. In this context, users have increasingly strong concerns about their sensible data stored, processed and travelling without having means to control their disclosure. The challenge consists then in providing each user with a control of his privacy-sensitive data and in guaranteeing the application of the appropriate privacy rules all along the lifecycle of these data. This should also allow the end-user to handle the famous privacy dilemma underlined by A. Westin [16]: "Each individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication of himself to others, in light of the environmental conditions and social norms set by the society in which he lives". Indeed, each end-user needs to self-position between the two following extremes:

- Widely expose his personal data in order to take benefit of personalized services, but taking risk regarding his privacy and
- Hide his personal data, but without benefiting from such complete personalized services.

IPTV Systems should support various privacy levels defining how personal data can be accessed and used.

In Table I, for every case study, a level of privacy is required. Therefore, the Privacy Control/setting module will translate the privacy policies required by the user into privacy information with respect to the service provider and to third parties.

A simple model could be adapted to some use cases according to the level of protection, as shown in Table I:

- **High privacy level**: no action authorized for the service provider (trace collection and automatic update) neither share information with third parties.

 ISBN: 978-1-61208-305-6

- **Average level**: service provider allowed to collect usage history and to automatically update preferences accordingly, but no sharing with third parties is allowed.
- **Low level:** service provider is allowed to collect usage history, to automatically update preferences accordingly and to share both of them (preferences and usage history) with third parties.

TABLE I: STUDY FOR DIFFERENT USE CASES PRIVACY LEVELS

Privacy Level / Use Case	My personal content follows me	Social media interaction	My TV content Away from Home (Nomadic)
High level			•
Average level		•	
Low level	•		

As shown before (in Section III. A), UP-TO-US adopts multi-identity mechanism and the privacy control will manage the divulgation of the information and profiles. The system handles different types of sensitive information like the name of the user, birth date, gender, specific location, contact information, history of visits, impairments, etc. Therefore, it is mandatory to implement a system to manage correctly the access to this information, assuring a determined level of data privacy chosen previously by the user and completed at the time after accessing the service.

IV. THE PROPOSED MODEL FOR PRIVACY

This proposed model facilitates identity exposure in nomadic situations. The question is how to choose the identity to expose.

We model the interaction between a user and a visited service provider as an extensive game [10], as show in Figure 7. An extensive game means that a user and a service provider take turns to make decisions and take actions.

We can model the extensive game as follows.

Two players: The user and the visited service provider.

A set of outcomes: in our example, there are the five outcomes: (Propose, Finish), (Propose, Abort), (Quit), (Accept, Abort), and (Accept, Finish).

Each player chooses and takes one action. The user and the service provider have their preferences, which will be represented by payoff values.

We believe that users want to maximize their privacy, while they access recommendation services. Service providers want to acquire more users' personal information while they provide services. We define their payoff functions as follows.

A user's payoff function, U, and a service provider's payoff function, S, at each node are, respectively:

$$U = Access - Exposure \quad (1)$$

$$S = Provide\ access + Exposure \quad (2)$$

From a user's perspective, "Access" brings him constant amount of benefit for a service, which is independent of his personal information exposure. The "exposure" component in (1) reduces the benefit as "exposure" increases. The service provider does not get enough identity information from a user and aborts. Then, the benefit is zero. As we can notice from (2), the more the user exposes, the more benefit a service provider receives.

Figure 7. User' strategy in inter domain situation

We use the nomadic access in hotel as an example to explain the game (inter domain access). Alice has a subscription and is in nomadic situation. She wants to access to her channels. Alice provides her pseudo name. The visited service provider asks for her name. Then, Alice makes a decision. After that, the visited service provider makes a decision: Alice may quit the service; she may accept the request and gives her name; or she may propose other options. Let's suppose that Alice proposes to provide only pseudo name.

Now, it is the service provider's turn to take decision. Suppose the service provider either aborts the checkout service or finishes the transaction. Therefore, there are five outcome cases. In case 1, Alice provides only her pseudo name, and the service provider finishes the transaction. In case 2, she provides only his pseudo name, and the service provider aborts the process. In case 3, she quits the checkout process. In case 4, she gives her name, and the visited service provider aborts the transaction. In case 5, she gives her name and the transaction finishes.

In Figure 8, Alice's preference is $\{1\} > \{5\} > \{3, 2\} > \{4\}$. That is, Alice prefers to get the same service by giving only his name.

 ISBN: 978-1-61208-305-6

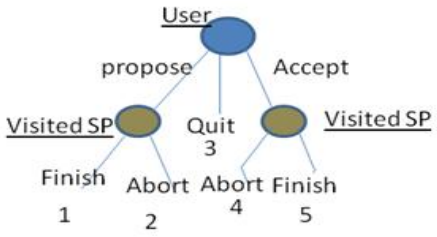


Figure 8. Extensive game between the user and the visited SP

She also prefers to get the service than not get it, even at the price of giving her name. Her least preferred outcome is that she gives her name without getting the service.

It is possible that two or more outcomes have the same preference. For example, {2} and {3} have the same preference for Alice. A subgame perfect nash equilibrium [17] is such that players' strategies constitute a nash equilibrium in every subgame of the original game.

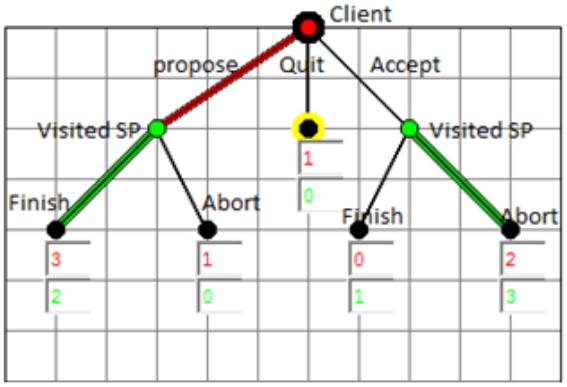


Figure 9. Online tool solution for the proposed game

This subgame is represented in red and green color in Figure 9. This subgame is unique and obtained by a Java tool [11] to affirm our analysis. In Figure 9, the payoff values for the user and the service provider in (*Propose*, *Finish*) = 3, *and* 2, respectively.

The analysis may be found by backward induction [10]. It is common method for determining subgame perfect equilibria.

To begin the process of backward induction, we assume that the client and the service provider will choose their preferences to reach an outcoume (determined by the payoff functions).

We must start from the bottom of the game and we walk through our example in Figure 9. The service provider may choose « *Abort* » or « *Finish* ». « *Abort* » gives the service provider a payoff of 0, where as « *Finish* » gives him a payoff of 2. Thus, he will choose « *Finish* ». Similarly, in the right, the service provider will choose « *Finish* » with a payoff of 3. He provides the service. In our example, the client may choose « *Propose* », « *Quit* », or « *Accept* ». From our previous discussion, the client's best choice is « *Propose* ».

The process continues until the root of the game is reached.

In our example, we have reached the root of the tree. So, the user chooses « *Propose* », and then the service provider chooses « *Finish* ». It is the subgame perfect equilibrium, the optimal choices for the client and the service provider.

V. CONCLUSION

This paper proposed a gateway to IPTV roaming services through nomadism. We presented the architecture of nomadic IPTV platform through UP-TO-US project. Moreover, we presented service personalization through EPG customization. This customization is depending mainly on four contexts (user, network, service and terminal). Also, we studied the privacy issue and its effects on service personalization. The sensitive information about users could reference the privacy-personalization phase in IPTV access especially in nomadic situations. For future work, we want to analyze more use the recommendation aspects using our proposed extensive game.

ACKNOWLODGMENT

This work has been supported by the Eureka Celtic UP-TO-US (User-Centric Personalized IPTV UbiquitOus and SecUre Services) European project. We would also like to thank Youness Oumzil for his help in the implementation.

REFERENCES

[1] UP-TO-US project: "https://up-to-us.rd.francetelecom.com/" [retrieved: July, 2013]

[2] Parks Associates Report "TV Everywhere: Growth, Solutions, and Strategies'', February 2011.

[3] Recommendation ITU-T Y.2091; "Terms and definitions for Next Generation Networks", February 2008.

[4] E. Abd-Elrahman and H. Afifi, " Moving to the Cloud: New Vision towards Collaborative Delivery for Open-IPTV", The 10 th International Conference on Networks ICN2011 in conjunction with GlobeNet 2011, St. Maarten,The Netherlands Antilles, 23-28 Jan 2011, pp.353-358.

[5] Recommendation ITU-T Y.1910; "IPTV functional architecture'', Appendix IV, Sept. 2008.

[6] Diameter Base Protocol, "http://tools.ietf.org/html/rfc3588" [retrieved: April, 2013]

[7] S. Song, H. Moustafa, and H. Afifi. "Personalized TV Service through Employing Context-Awareness in IPTV/IMS Architecture,'' In Proc. of the 3rd International Conference on Future Multimedia Networking (FMN'10), Krakow, Poland, LNCS, volume 6157, June.2010, pp.75–86. Springer-Verlag,

 ISBN: 978-1-61208-305-6

[8] S. Song, H. Moustafa, and A. Afifi, "A survey on personalized TV and NGN services through context-awareness". ACM computing surveys, January. 2012, vol. 11, n° 1, article 4

[9] S. Song, H. Moustafa, and A. Afifi, "Enriched IPTV services personalization," ICC '12: IEEE International Conference on Communications, 10-15 June . 2012, Ottawa, Canada, pp. 1934-1939.

[10] M. J. Osborne, " An introduction to game theory;" Oxford University Press, August 2003

[11] Extensive Form Game Solver: "http://www.gametheory.net/Mike/applets/ExtensiveForm/ExtensiveForm.html" [retrieved: July, 2013]

[12] A. Elmisery, and D. Botvich, "Privacy aware recommnder service for IPTV networks," FTRA' 11 International conference on multimedia and ubiquitous engineering, June. 2011, pp. 160 - 166

[13] P. Shoval, V. Maidel, and B. Shapira, "An Ontology- Content-Based Filtering Method," In I.Tech-2007 - Information Research and Applications, 2007.

[14] R. Pampapathi, B. Mirkin, and M. Levene, "A Review of the Technologies and Methods in Profiling and Profile Classification," EPALS Technical Report, April, 2005.

[15] I. Cantador, A. Bellogín, and P. Castells, "A Multilayer Ontology-Based Hybrid Recommendation Model," AI Communications, Special Issue on Recommender Systems, IOS Press 2008, pp. 21(2-3) 203–210.

[16] A. Westin, "Privacy and freedom," Fifth ed., New York, U.S.A.: Atheneum, 1968.

[17] I. Obara, "Subgame Perfect Equilibrium," UCLA, 2012

 ISBN: 978-1-61208-305-6

Toward a Global File Popularity Estimation in Unstructured P2P Networks

Manel Seddiki*, Mahfoud Benchaiba‡
*,‡ University of Sciences and Technology Houari Boumediene
Computer Science Department, LSI laboratory
Algiers, Algeria
* e-mail: sed.manel@gmail.com
‡ e-mail: benchaiba@lsi-usthb.dz

***Abstract*—In unstructured P2P networks, replicating most popular files is one of mechanisms, which improve file lookup performances, such as lookup delay and success rate. However, measuring global file popularity is a challenging task because this estimation must consider requests of all peers for this file whereas in unstructured P2P networks like Gnutella, the peer has no global view of the network. Some researches have been done to measure this parameter. Nevertheless, this estimation is still away from reality because the peer, which calculates file popularity, doesn't consider file popularity estimations of the other peers. In this paper, we try to define a way to calculate a global file popularity based on local estimation of the peer and estimations done by the other peers participating in the network. Our first simulation results reinforce our theoretical formulas and show that our measurement is closer to the real one. More details will be provided and simulation tests will be added in our future contributions.**

***Keywords*—Unstructured P2P networks, global file popularity, file lookup,request packets, replication.**

I. INTRODUCTION

Peer-to-peer (or P2P) networks came to replace client/server systems and were developed over Internet in recent years. The basic idea of P2P is to link users in order to exchange information without using any intermediate server. Thus, P2P network is a distributed system of interconnected peers, which are both clients and servers. The P2P paradigm was firstly used for file-sharing applications such as Napster [1] and Gnutella [3], which allow users to lookup, share and download files.

Napster uses a server which indexes all the information about peers and their files. If a peer wants to lookup for a file, it sends a request to the server, which connects it directly with peers storing this file. The server facilitates the lookup procedure and improves the lookup latency, but it is the weakness of the system because if it breaks down, the whole system stops. Gnutella came after Napster and erased centralization idea. Indeed, Gnutella works on an unstructured P2P network architecture, where there is no server and each peer must know the other peers participating in the P2P network and their shared content by itself. A peer wishing to lookup for a shared content, such as a file, broadcasts its request to all its neighbors, which do the same with their neighbors until the file is found or the Time To Life (TTL) expires. This technique is denoted as flooding [3]. However, the flooding main drawback is the high overhead that causes a scalability issue. Many alternatives to flooding have been proposed to make file lookup technique more efficient, such as using probability based on previous lookup results ([4] and [5]), using progressive TTL called Expending Ring such as [6] or using Random walk technique such as [7].

Another way to improve file lookup performances in P2P unstructured networks is replication, as presented in [8], [9], [10], and [11], which consists in the replication of most popular files in other peers to ensure their availability, increase lookup success rate and decrease lookup hops and delay. Performances of these replication strategies depend on the popularity parameter precision. Indeed, the closer is the popularity estimation from reality, the better is the replication strategy performance. As a consequence and for our point of view, the file popularity measurement in such replication strategies is then crucial to decide which files have to be replicated. However, most of these strategies don't focus on this measurement and briefly define file popularity calculation based only on local estimation of the peer. This is maybe due to the fact that in P2P unstructured architectures, the peer is blind and has no global view of the network and this makes global popularity estimation a challenging task. In this paper, we focus completely on this issue and try to define the file popularity notion and four evident criteria that the file popularity estimation must respect. After that, we propose a way to calculate the file popularity according to and respecting those creteria. This calculation is based both on local estimation of the peer and estimations done by the other peers participating in the network. Indeed, considering the estimations of the other peers allows having a global-like estimation of the popularity which is closer to the reality than the local estimation.

This paper is organized as the following: In Section II, we introduce some interesting researches which calculate file popularity used in variety of contexts, such as content replication strategies and file lookup enhancement. In Section III, we describe our approach in detaills. We begin first by describing the P2P network architecture and environment that we consider in our approach then, we describe our file cache structures and define the popularity notion according to our point of view. After that, we explain our file popularity measurement and finally, we discuss some points. In Section IV, we introduce simulation environnement, describe the different simulation tests and compare our estimated popularity with the real popularity. In the end of this paper, we give a brief summary of this paper's content and next contributions to finalize our

 ISBN: 978-1-61208-305-6

work.

II. Related work

Despite of the file popularity importance in the replication and file lookup area, there are no consistent investigations in calculating a global file popularity, which is close to the real popularity. However, many replication strategies, such as those in [8], [9], [10], and [11] proposed some simple popularity measurments. In [9], the file popularity measurement is simply obtained by counting the number of accesses of each file f as follows: Peer P2 asks for a file f from the peer P1 ; P1 is able to provide file f or its index; P2 accesses P1 to retrieve file f ; P1 increments f popularity as follows:

$$P_f = P_f + 1 \tag{1}$$

In [8], the Q-replication strategy defines a popular file as a file which is frequently accessed. Each peer maintains a table containing the file name and the file popularity. The file popularity for each file f is calculated as follows:

$$P_f(t+1) = P_f(t) + \eta \frac{R_f(t)}{N(t)} * 100 \tag{2}$$

$R_f(t)$ is the number of requests seen by the peer for the file f at time t , N(t) is the total number of requests received by the peer at time t and η is a constant variable. The popularity is updated according to (2) after a fixed total number of requests received.

Another way to estimate file popularity is described in [10]. In this paper, the popularity is defined as the request rate for a file f and it is calculated as follows:

$$P_f = \frac{R_f}{T} \tag{3}$$

$R_f$is the number of requests peer have seen for the file f and T is the amount of time the peer has been up. The popularity is updated each time the peer receives a request for file f.

In [11], a dynamic data replication strategy is propsed. Indeed, to improve grid system performances, authors propose a dynamic strategy to replicate data in several sites of the grid considering crash failures in the system. The strategy is based on 2 parameters: Availibility and popularity of data. The popularity of the data f is calculated in this paper as follows:

$$P_f = \frac{R_f}{N} \tag{4}$$

$R_f(t)$ is the number of requests demanding f and N is the total number of all requests.

In our opinion, file popularity estimation has to respect four criterions:

- The popularity value is a rate and must be between 0 and 1.
- As the popularity depends on external actors (in our case, file requests), it must increase when request rate for this file is high and decrease when it is low. Let us take for example an artist-painter: His popularity depends on its fans (external actors) , it increases when its fans request highly its paintings and it decreases when not.
- Popularity value must be influenced explicitly or implicitly by time and this criterion is related to the previous point.
- Popularity must be based on global knowledge of requests circulating in the network.

All of [8], [9], [10], and [11] are based only on local estimations of the peer in popularity measurement. They don't acquire a global knowledge about the file popularity. In [8] and [9] and according to (1) and (2), the popularity measurement is cumulative, which means that the value will never decrease. Moreover, it is not between 0 and 1. In [8] and according to (1), popularity is not influenced by time and in [10], the popularity is defined as the number of requests for the file f by time unit. This leads to simply request rate and not popularity estimation. We conclude that criteria mentionned above are not all respected by [8], [9], [10], and [11]. In this paper, we try to consider all those criteria to provide a file popularity definition and calculate its estimation in order to make it close to the real value.

III. Our contribution

In this section, we describe our file popularity measurment which is based on both local popularity measurment of the node and popularity measurement of its neighbors. The idea is to have a global-like knowledge about the file by using neighbors which did the same with their neighbors and so on.

A. P2P network environement

We consider unstructured P2P architecture where peers index their own files and have no knowledge about the other shared files in the network and their locations at the beginning. Our file popularity measurement operates during file lookup phase and each peer is supposed to have at minimum, one neighbor.

B. Files cache

Each peer X participating in the P2P network maintains 2 structures denoted by S1 and S2 as shown in Figure.1. The first structure S1 stores local files and files discovered from file request packets passed through X. Each entry of S1 contains information that the peer knows about the file which, are the file key, number of requests passed through X for this file, local popularity and global popularity calculated by X. All explanation about how to calculate local and global popularity will be given in next section. The second structure S2 stores all the files's popularities of X's neighbors. S1 and S2 will be used to extract all necessary information needed in the computation of file popularity. S1 is initialized by adding local files of peer X with number of requests=0, local popularity=0 and global popularity=0. New entries in S1 are added when the peer X discovers new information about a file in the request packet passed through it and increment number of request by 1 for the concerned file. Moreover, peer X exchanges periodically its file list with its neighbors. S2 is initialized and updated when X receives this list.

 ISBN: 978-1-61208-305-6

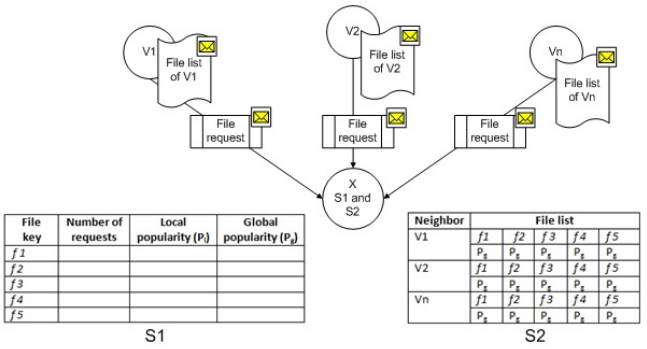


Figure. 1: File cache structure

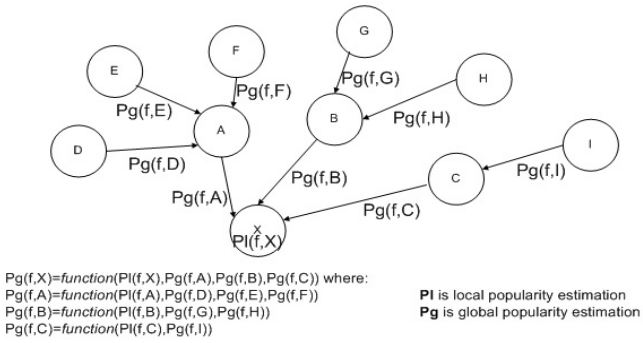


Figure. 2: Global popularity estimation scheme for peer X

C. File popularity definition

For the best of our knowledge, a file is popular if it is highly requested in the network. Several definitions of file popularity have been disscussed in the related work section. We define the file popularity as the ratio between the number of requests for the file f and total number of requests in the entire network formulated as follows:

$$P(f,X) = \frac{number\,of\,requests\,for\,the\,file\,f}{total\,number\,of\,all\,requests} \quad (5)$$

The real file popularity estimation presented in (4) can be only calculated by a global observer, which has a global view of the entire P2P network. However, peers have no global view in unstructured P2P network. Indeed, each peer is blind and has only local knowledge about the file. This local information is not enough to have a real estimation of file popularity. Thus, our goal is to find a way to bring global-like information about files and include it with local information to have file popularity estimation closer to the real estimation. In our approach, peers benefit from the knowledge of the other participating peers through neighbors. In fact, the peer calculates file popularity based on its own knowledge and knowledge of its neighbors. Knowledge of neighbors is obtained based on the own knowledge of neighbors and the one of their neighbors, and so on, as it shown in Figure. 2 . In this way, all peers cooperate to provide a global view of the file in the network and thus, estimate a file popularity closer to the real one.

D. File popularity estimation

In this section, we define our file popularity measurement. It is composed of two major steps. The first step is the local popularity estimation, which is based on the local knowledge of the peer about the file. Local knowledge is obtained by exploiting file request packets passed through the peer. The second step is global popularity estimation, which is based on the local popularity estimated in the first step and the global popularity estimated by direct neighbors. At the beginning, the local and global file popularities are defined by 0 for each file *f*. Local file popularity is updated when the peer receives a request packet form its neighbors and global file popularity is updated when the peer receives a file list from its neighbors.

1) Local file popularity estimation: It consists on calculating file popularity based on local knowledge of the peer. The local popularity of the file f for the peer X denoted by $P_l(f,X)$ is defined as the ratio of known requests for the file f denoted by R_f to all known requests denoted by R. It is formulated as follows:

$$P_l(f,X) = \frac{R_f}{R} \quad (6)$$

R and R_f are obtained from structure S1. The local popularity is updated each time a peer receives a request packet.

2) Global file popularity estimation: Local popularity estimation is not enough to reflect the real value. Indeed, we need to have a global estimation of f's popularity by considering both the local estimation formulated in (5) and global estimations of neighbors. It is calculated as follows:

$$P_g(f,X) = \frac{(P_l(f,X) + \sum_{j=1}^{|V|} P_g(f,V_j)}{(|V|+1)} \quad (7)$$

Where $|V|$ is number of neighbors of peer X which, have calculated global popularity of file f , $P_l(f,X)$ is local popularity of the file f for the peer X and $P_g(f,V_j)$ is global popularity of the file f for the neighbor V_j such as $1 \leq j \leq |V|$. $|V|$ and $P_g(f,V_j)$ are obtained from S2.

E. Discussion

The global popularity is calculated when the peer receives file list from its neighbors. This calculation is done in two ways:

- The first way (which, we consider in this paper) is the periodic list reception from neighbors. In this case, the challenge is to select the suitable delay because if it is too small, the overhead increases in the network due to high exchange of file lists and if this delay it is too large, this may result in imprecision on popularity estimation and lack of updates concerning file requests and peers disconnections.
- The second way is the on-demand list reception, which means that if the peer wants to calculate file popularity for replication or for other purpose, it requests its neighbors for the file list. The advantage on asking for file list on-demand is that neighbors send only the

 ISBN: 978-1-61208-305-6

TABLE I: Simulation parameter

Simulation time	1000s
Average neighbors	3
Number of nodes	100
File list delay	40s
Node join and departure	Lifetime churn with lifetime=600s
request load	1 request for random file per 60s

concerned file and not all files and this will decrease file list size but the drawback is the time wasted on waiting for the file information to be received.

Our popularity estimation is based both on local knowledge of the peer and global-like knowledge acquired through neighbors as explained in previous sections. This estimation is bounded by 0 and 1. Moreover, it increases when request number for the concerned file is high comparing with the other requests and decreases when not and time influences the estimated value implicitly through those requests. Hence, the four criterions are respected.

IV. Simulation

In order to compare our file popularity estimation with the real popularity value, we implemented our file popularity algorithm and a global observer algorithm on OverSim [12] with Omnet++ [2]. The P2P network is composed of 100 peers, which may join and leave according to lifeTimeChurn=600s as shown in table I. Each peer in the network has a random number of local files limited to 100 maximum and enriches its structures S1 and S2 through file list exchanged between neighbors and request packets passed through the peer. A peer sends a request for a file choosen rondomly every 60s. We chose one file with key=E88 from the network to observe its popularity evolution. Our initial results are obtained by comparing our popularity estimations with the global observer estimations. In Figure.3, the thick line with square symbols represents real file popularity evolution calculated by the global observer and the other thin lines represent file popularity estimated by some peers participating on the network according to our approach. Thus, Figure.3 shows that all peers estimate popularity values that match closely with the real popularity calculated by the global observer. This is due to the cooperation between all peers in order to allow having to each single peer, a global-like view of the file. A best view of this match is represented in Figure.4 where the general bahaviour of the system represented with triangle symbols match closely with the global observer behaviour represented with square symbols. These simulation results reinforce our theoretical formulas and prove that our file popularity estimation is efficient in an unstructured P2P network.

V. Conclusion

Estimating real file popularity in unstructured P2P networks is a hard task because peers are blind and have no global view of the network resources. In our point of view, calculating file popularity value, which is close to reality must respect four criterions : It must be bounded by 0 and 1; it must increase and decrease according to external actors; it must be influenced by time implicitly or explicitly; it must be based on a global-like knowledge about the concerned file. Several researches proposed to measure the file popularity, but not all the four criteria were considered.

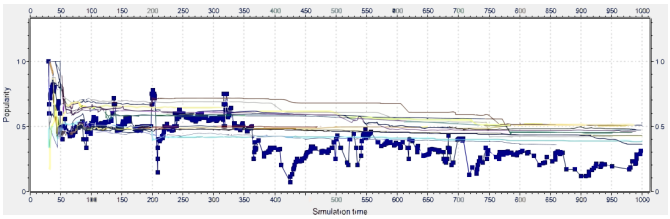


Figure. 3: Real popularity vs estimated popularity of file E88

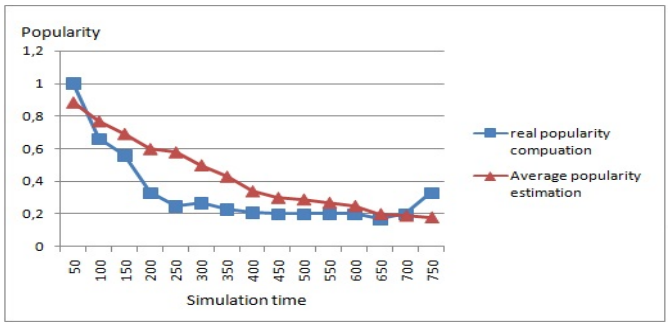


Figure. 4: Real popularity vs estimated popularity of file E88

In this paper, we define file popularity and we propose a measurement for it respecting the four criteria. Our first simulation results reinforce our theoretical formulas and show that our measurement matches closely with the real one. These initial results prompt us to investigate more about this rate. More details will be provided and simulation tests will be added, such as the impact of the search rate on the popularity deviation in our future contributions.

References

[1] (1999) Napster. [retrieved:march , 2013]. [Online]. Available: http://www.napster.co.uk.

[2] (2002) Napster. [retrieved:april , 2013]. [Online]. Available: http://www.omnetpp.org.

[3] M. Ripeanu and I. Foster, "Mapping the gnutella network, Internet Computing," IEEE, vol. 6, 2002, pp. 5057.

[4] R. Gaeta and M.Sereno, "Generalized probabilistic flooding in unstructured peer-to-peer networks, Parallel and Distributed Systems," IEEE Transactions, vol. 22, December. 2011, pp. 2055 2062.

[5] S. Margariti, "A novel probabilistic flooding strategy for unstructured peer-to-peer networks, 15th Panhellenic Conference, " September. 2011, pp. 149153.

[6] Q. Lv, P. Cao, E. Cohen, K. Li and S. Shenker, "Search and replication in unstructured peer-to-peer networks, Proceedings of the International Conference on Supercomputing," June 2002, pp.22-26.

[7] C. Gkantsidis and A. Saberi, "Random walks in peer-to-peer networks, Proc. of IEEE INFO-COM," vol. 1, March. 2004, pp. 130140.

[8] SM.Thampi and K. Sekaran, "Review of replication schemes for unstructured P2P networks,"' arXiv preprint arXiv:0903.1734, no. March. 2009, pp. 67.

[9] S. Mohammadi, H. Pedram, and A. Farrokhian, "An enhanced data replication method in p2p systems, Journal Of Computing," vol. 2, November 2010, pp. 7882 .

[10] J. Kangasharju, W. Ross, and D. Turner, "Optimal content replication in p2p communities, Manuscript" 2002.

 ISBN: 978-1-61208-305-6

[11] B. Meroufel and G. Belalem, ”Dynamic replication based on availability and popularity in the presence of failures, Journal of Information Processing Systems,” Vol.8. June. 2012, pp. 263278.

[12] I. Baumgart and S. Krause, ”Oversim: A flexible overlay network simulation framework, 2007 IEEE Global Internet Symposium,” May. 2007, pp. 7984.

 ISBN: 978-1-61208-305-6

Exploiting Semantic Indexing Images for Emergence Recommendation Semantics System

Damien E. ZOMAHOUN
University of Burgundy
Dijon, FRANCE
Email: damien_zomahoun@etu.u-bourgogne.fr

Pélagie Y. HOUNGUE
University of Burgundy
Dijon, FRANCE
Email: pelagie.houngue@u-bourgogne.fr

Kokou YETONGNON
University of Burgundy
Dijon FRANCE
Email: kokou@u-bourgogne.fr

Abstract – Thanks to the efforts of the Semantic Web Community (W3C), images can be semantically indexed with metadata. The explicit representation of image contents is made possible by using ontologies that provide a common and shared understanding of a domain at both human users and application levels. The approach that we are proposing in this paper is a semantic indexing of images based on conceptual method. To make efficient the semantic indexing, we also propose a recommender system. User profiles: static and dynamic profiles are combined and supported by the system we have developed to suggest recommendations to users. The preferences of each user are taking into account to provide customized recommendations.

Keywords- image; semantics; ontology; indexing; recommendation

I. INTRODUCTION

The growth of multimedia data and in particular images caused not only a need for storage but also the need to get access to those images. So as to these data to be usable, they need to be effectively referred back to as in a catalogue. The techniques presented below, called indexing, propose to attach to an image a set of descriptors that describe their content.

Many approaches seek the use of semantics to extract the representative of images content descriptors. These descriptors are then used to allow the system to retrieve the images of interest to the user. A set of keywords, names, nominal sentences are mapped to the concepts that they represent [2]. In these approaches, an image is represented as a set of concepts. To achieve this, the semantic structures of image representations are needed. These structures can be dictionaries, taxonomies or ontologies [3]. They can be either manually or automatically generated. They are widely used to improve the efficiency of images retrieval. There are generally three types of indexing: classic, conceptual and ontological. The classical indexing is based either on lexical or syntaxical analysis of the images content by taking into account keywords occurrences. The conceptual indexing is a statistical approach that aims to extract the semantics contained in the images. This approach groups terms that have common features in images and considers that each group represents a semantic. The terms chosen should allow to retrieve the relevant images with respect to the representation of user needs. Two parameters are taken into account in classical indexing: language of representation and discriminating power. [4][5][6] [7].

A new generation of methods is to consider the concepts rather than words. The conceptual indexing allows to identify the concepts and / or instances of ontology that appear in the images. The approach proposed in [15] aims to understand the specific requirements of users in order to meet their needs. Users propose instances of concept in their queries such as the acquisition time and the type of sensor and especially select the concepts by which they are interested. The researchers also assumed that the terms can carry a semantic structure whose they try to extract the concepts as a unit of semantic. To achieve this purpose, several approaches have emerged. The approach proposed in [7][8][9] aims to avoid the polysemy and synonymy of terms used as descriptors by conventional statistical approaches. It groups terms with common characteristics in their appearance in the images. Another approach consists of identify the elements of the ontology. This approach was used in [10] [11][12][13]. Other approaches extract “expressions”. The extraction of expressions is important because the instances of the concepts are often composed of such elements [10][14].

Another type of ontological indexing approach is to rely on ontologies to retrieve images; this type of indexing is called ontological indexing. The ontological indexing put forward the fact that the meaning of textual information depends on the conceptual relationships between the objects to which they refer to [1] [16]. Ontological indexing is possible only by the existence and use of resources explicitly describing the information corresponding to objects [17][18]. Regarding ontology usage for images indexing, Khan [19] proposed a method based on sub-trees "regions" of ontology. Regions of an ontology represent different concepts. Concepts that appear in a given region are mutually disjoint concepts from other regions. The region containing the largest number of concepts is selected. Then all the selected concepts that also appear in other regions are deleted. In a region, the selection is made through the use of "semantic distances", by taking into account of paths between

 ISBN: 978-1-61208-305-6

concepts in the ontology. The concepts that correlate with the greatest number of other concepts are selected for indexing. Woods [20] proposed the same method of indexing, but his approach retrieves similar images.

The studies performed in this paper consist in comparing the existing semantics indexing methods through large collections of images and present their advantages. Following, we will propose a technique that meets better the needs of users. Finally, we will propose a recommendation system for the semantics used by users to retrieve images. This includes a new factor to better take into account the user interactions with the retrieval system to perform specific recommendations.

II. EMERGSEM APPROACH

We note that once the images are annotated, different needs for access to these images appeared, each corresponding to a specific action: sequential scan of a set of images in the case where the user does not really have idea of what he wants, image search, when the user knows exactly what he wants and finally the image classification, which helps users to combine images with similar features and thus provides a simplified representation of an ordered set of images.

The architecture of the suggested system is shown in Figure 1.

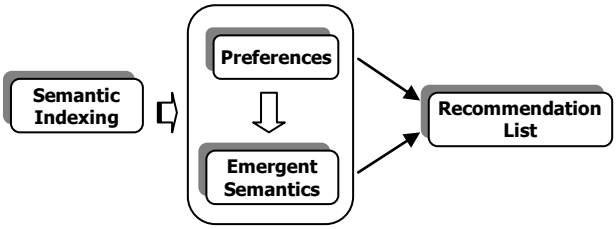


Figure 1. EMERGSEM model

We propose the next two steps to realize EMERGSEM system. The system is equipped with image retrieval functionality and recommendations ranking technique to facilitate image retrieval and recommendations generation. The approach proposes a technique based on ontology concepts. Concepts are used to select image semantics that are then used to retrieve images. Then, the system makes semantics recommendation using three fundamental steps: acquiring preferences from the user's input data, computing recommendations and suggesting the recommendation to the user.

III. SEMANTIC INDEXING

In this section, we present an approach for semantic indexing of images based on concepts. To facilitate semantic indexing of images, we propose a conceptual indexing to end users. So users can provide keywords that represent instances of concepts of ontologies used to store the semantics of images they want as we show in Figure 2.

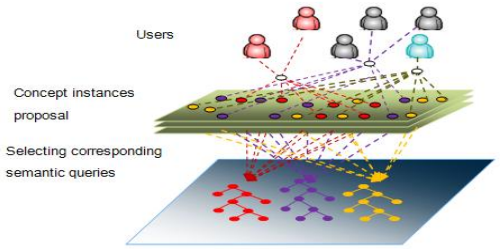


Figure 2. Semantic indexing

The system selects the most similar semantics corresponding to instances proposed by user i.e., the concepts belonging to semantic of the retrieved image. This selection is based on the following algorithm.

InstanceConceptSem (k_i1) = first instance of concept retrieved in semantic i
. . .
InstanceConceptSem (k_in) = n instances of concept retrieved in semantic i
Instance k
List list=similar retrieved semantics
list [0] =One semantic
For eachsemantic==i do
Prob [i] = (InstanceConceptSem (ki1) +...+ InstanceConceptSem (kin))/Nc
End for

Figure 3. Retrieval Algorithm

For ki instances submitted by the user, we define the probability that these instances appear in the instances of each semantic displayed Nc. The retrieval probability is then computed as follows:

$$\text{Prob} = \frac{\sum \text{ki}}{\text{Nc}} \quad (1)$$

Based on the semantic interpretations provided, the similar semantics containing these instances can be obtained. Lastly, the image semantics that contain greater instances are retrieved. Subsequently, the user can select the appropriate semantic of semantics displayed to index the image.

The process of the concept-based image retrieval depicted can be described as follows:

(1) User proposes keywords indicating the content of images.

(2) EMERGSEM system verifies the presence of the instances proposed by user in each concept of the ontologies.

(3) Once the instances are validated, the program invokes an ontology query service. The semantics containing the instances are displayed thanks to the ontology search engine.

(4) User selects the semantic that describes better the needs image. Image is retrieved from the image database.

(5) Finally, the results are displayed to the user.

IV. RECOMMENDATION SYSTEM

Recommendation system has been a hot research topic in recent years. To recommend items to a user, the system must have a representative profile preference. To build this, it must collect

 ISBN: 978-1-61208-305-6

information about it, either directly or indirectly [21][22]. Recommender systems help users to manage information overload by providing personalized advice on content and online services. The term "recommendation system" generally describes a system that produces customized recommendations to users, and has the effect of leading the user to interesting items in a large space of possible options [23][25].

A recommendation system we propose aims to recommend images semantics to a user in correspondence with its tastes and preferences. The aim is both to minimize the time spent on research, but also to suggest relevant semantics that would not be spontaneously consulted and increase the overall satisfaction of users.

The first step in the realization of the recommendation system is to extract the profiles of users. The next section focuses on this purpose.

A. Acquiring Profiles

Acquisition of user profiles is composed of two important steps: extraction of static and dynamic profiles.

The static profiles are also called independent part of the domain. They take into account any data that has no connection with the domain. There may be personal user information such as professional status. This part does not require large resources since users, before using the system are required to create an account and thus to provide such personal information.

The dynamic profiles are also called dependent part of the domain or the active model. It consists of data that represent the needs, interests and goals of the user i.e., user preferences. This part will be constructed by the system in response to user interactions with the system: a history of the user's interactions with the recommendation system. To achieve this, the system needs to collect such data on assessments of the user. The analysis of these data is then used to build a model of the user's preferences that will be used by the system to recommend the semantics deemed relevant for the user. A model of the user's preferences, i.e., a description of the types of semantic that interest the user is represented. There are many possible alternative representations of this description, but one common representation is a function that for any semantic predicts the likelihood that the user is interested in that semantic. For efficiency purposes, this function may be used to retrieve the n semantics most likely to be of interest to the user [26].

B. Classification of Preferences

What we learned from this work is that the comparison between the profiles leads to the formation of user groups close to each other, groups called "communities". So we can say that the notion of community is a key factor in a recommender system to produce recommendations [24]. It is clear that the positioning of users in the spaces depends crucially on the dynamic profiles. The dynamic profiles of each user then evolve along with the user himself.

To group profiles, we think that the formal concept analysis [29][30][31] and Galois lattices [32][33][34] will be indispensable. A lattice of Galois can regroup, exhaustively objects in classes, called "formal concept", using their shared properties. A lattice is typically based on a Boolean matrix, called matrix context and denoted C, whose rows represent a set of objects O that we wish to describe and columns, a set of attributes A that these objects have or have not.

TABLE I. FORMAL CONCEPT OF SEMANTICS

	Static Profiles			
Dynamic Profiles	User 1	User 2	User 3	User 4
Semantic 1	X			X
Semantic 2		X	X	
Semantic 3	X	X	X	
Semantic 4				X
Semantic 5			X	
Semantic 6		X		
Semantic 7		X		X

Suppose we have a description of the following profiles (see Table I). This description is based on the list of dynamic profiles that users have chosen or not. Possession of the property $a \in A$ by object $o \in O$ reflects the existence of a relationship I between them: aIo. The existence of this relation I between O and A is materialized in the matrix of context C by a value "true" or "false". The triplet $K = (O, A, I)$ is called a formal context or context. A set X ⊂ O is the set of attributes jointly owned by all object X and is given by the function f:

$$f(X) = \{a \in A | \forall o \in X, oIa\} \quad (2)$$

Inversely a set Y ⊂ A is the set of objects jointly owned by all object Y and is given by the function g: $g(Y) = \{o \in O | \forall a \in Y, oIa\}$ (3)

The pair (f, g) is called a Galois connection. A concept is any pair C = (X, Y) ⊂ O × A, such that objects in X are the only one to have attributes in Y; in other words X x Y is, if we add permutations of O and A, a maximal rectangle in C, i.e.

$$f(X) = Y \,\&\, g(Y) = X. \quad (4)$$

To illustrate this approach of formal Concept, the Table I shows that the set X= {Semantic 2, Semantic 3} gives one formal concept since $f(X) = \{\text{User 2, User 3}\} = Y$ et $g(Y) = X$, and this formal concept is ({Semantic 2, Semantic 3}, {User 2, User 3}), while the set X' = {Semantic 1, Semantic 4} doesn't give a formal concept because $f(X') = \{\text{User 4}\} = Y'$et $g(Y') =$ {Semantic 1, Semantic 4, Semantic 7}

The Galois lattice is represented by a Hasse diagram as shown in figures 4. "Sem" means Semantic.

 ISBN: 978-1-61208-305-6

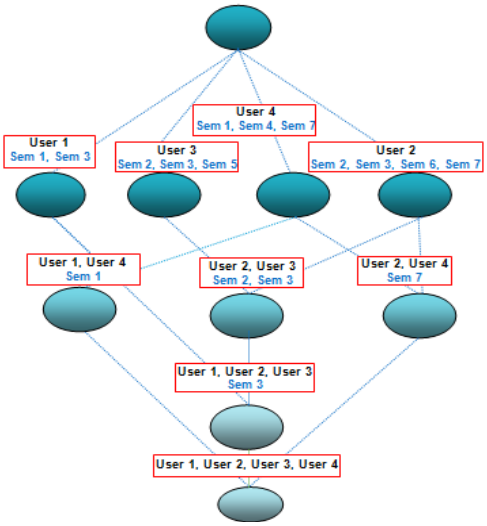


Figure 4: Display Galois lattice of users profiles

In this Figure, the static profiles are connected to dynamic profiles. For example we see that the users 1, 2 and 3 have selected semantics 3 while the users 2 and 4 chose the semantic 7.

One of important advantages of classification based on Galois lattice is that for a given formal context table the resulting lattice is unique, and it is exhaustive. This classification will allow us to find all the groups of static profiles in relation with a group of dynamic profiles and represent them similarly.

C. Emergent Semantic

Image semantics are provided to users after the proposition of instances. Users choose appropriate semantic in the list, i.e., the semantic that better meet their research needs. Once a semantic is used to search for images, a weight is assigned by the recommendation system which is responsible for the link between the semantic and the user. The users interact with the system, and the semantics used will be evaluated by the system. The system can therefore recommend these semantic to them. The weight is calculated by:

$$\mathrm{Wht}_{i,k} = \sum \mathrm{Pbt}_{i,k}, \quad (5)$$

where Pbt is the probability of a semantic i to be chosen in image k.

D. Recommendation List

Recommendation list may then be introduced, once user profiles are grouped. It is to look for similarities between the dynamic profiles of each constituted group to make customized recommendations. Similarity measures considered here satisfy the following properties for all (u, v) $\in U$:

- sim (u, v) $\in [0; 1]$; (sim mean similarity)
- sim (u, v) = 1; if and only if u and v have the same common profile;
- sim (u, v) = 0; If u and v do not have a elements of comparison [24].

The method chosen to determine the user's profiles similarity is cosine similarity since the cosine similarity seems very promising. It provides an accurate measure of similarity [27][28].

The recommendation list given to user is consisted of two parameters: the personalized recommendations representing the preferences of each user and the general recommendations representing a mostly used semantic of each image, that are the emergent semantics. Unlike specific recommendations the emergent semantic is recommended to all users.

Let u_1 and u_2 be two users with dynamic profiles specifying their utility functions of the subsets $I_1 \subseteq I$ *and* $I_2 \subseteq I$. We then calculate the similarity typically using for example the cosine similarity [25] by:

$$cos_{ut}(\mathrm{ut}_{u_1}, \mathrm{ut}_{u_2}) = \frac{\sum_{i \epsilon I_1 \cap I_2} \mathrm{ut}_{u_1}(i).\ \mathrm{ut}_{u_2}(i)}{\sqrt{\sum_{i \epsilon I_1} \mathrm{ut}_{u_1}(i)^2}.\sqrt{\sum_{i \epsilon I_2} \mathrm{ut}_{u_2}(i)^2}} \quad (6)$$

If $cos_{ut}(\mathrm{ut}_{u1}, \mathrm{ut}_{u2}) \geq$ Threshold then u_1 and u_2 are similar. Note that the threshold varies from one image to another, and this threshold is not stable. In order to compute the threshold between the instances of concepts, we calculate the average of common with proposed instances. Let x and y denote the feature sets of the common and proposed instances, respectively for image I. The threshold is,

$$\text{Threshold} = \frac{\sum x_I}{\sum y_I}, \quad (7)$$

For example for image 1 (Table III, Table IV), the threshold= 15/21= 0,714.

V. EXPERIMENTATION

To illustrate how the combined approaches perform in practice, it was evaluated on a real-world semantics recommendation application and compared its performance with the simple semantics indexing. The following table (Table II) gives the variations of images indexing durations. We note that the data in the first column of the table (T_i (hour)) are higher than those of other columns which remain stable.

The fundamental reason is that the first index is performed without any semantic recommendation. Users have suggested instances of concepts to retrieve images. But in the other columns, the search time is greatly reduced because the system took into account the preferences of the users to make their recommendation. The result is that they have a huge time saver. The following figure (Figure 5) shows indexing duration variation.

TABLE II. EXAMPLE OF INDEXING DURATION

	T_i (hour)	**T_{i+1} (hour)**	**T_{i+2} (hour)**	**T_{i+3} (hour)**
Image 1	0,0500	0,0041	0,0033	0,0032
Image 2	0,0417	0,0027	0,0029	0,0030
Image 3	0,0672	0,0028	0,0031	0,0033
Image 4	0,0375	0,0032	0,0032	0,0029

 ISBN: 978-1-61208-305-6

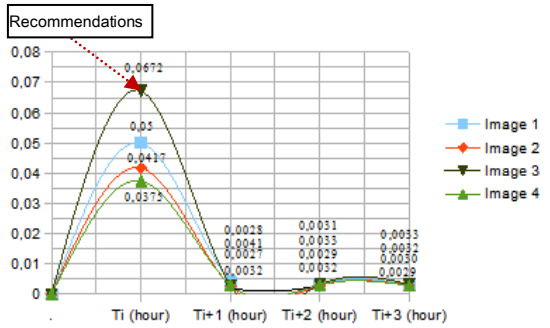


Figure 5: Indexing duration

The figure 5 shows a graph with four curves representing the indexing duration of the 4 images. We note that the curves decrease when users had the recommended semantics for images semantic indexing. Then, the curves remained stable with values around 10.8 seconds. The recommendation system has helped to save time during semantic indexing.

The following tables (Table III, IV and V) shows the different steps of calculating the similarity between user profiles to make them recommendations based on their preferences. Table 3 is an example of instances of concepts proposed by users.

TABLE III. CALCULATION OF INSTANCES PROPOSED BY SEMANTIC
U means User, and Sem means Semantic

	Instances proposed								
	U_1			U_2			U_3		
	Sem 1	Sem 2	Sem 3	Sem 1	Sem 2	Sem 3	Sem 1	Sem 2	Sem 3
Image 1	08	-	-	-	07	-	06	-	-
Image 2	-	06	-	10	-	-	-	07	-
Image 3	-	-	08		-	09	11	-	-
Image 4	05	-	-	-	04	04	-	-	-

TABLE IV. IDENTICAL INSTANCES OF CONCEPTS BETWEEN USERS

	$U_i, U_{j(i \neq j)}$		
	$U_1 \cap U_2$	$U_1 \cap U_3$	$U_2 \cap U_3$
Image 1	04	07	04
Image 2	05	02	08
Image 3	07	04	06
Image 4	02	02	04

TABLE V. DYNAMICS PROFILES SIMILARITY BASED ON COSINE

	$cos_{ut}(\mathbf{ut}_{u_1}, \mathbf{ut}_{u_2})$		
	$U_1 \cap U_2$	$U_1 \cap U_3$	$U_2 \cap U_3$
Image 1	0,2539	0,8750	0,3809
Image 2	0,4167	0,0952	0,9142
Image 3	0,6805	0,1818	0,3636
Image 4	0,2000	0,1142	0,5714

We note that for the image 1, user 1 proposes 8 instances of semantic 1 while user 2 provides 7 instances of semantic 2 and user 3 has 6 instances of semantic 1, etc. Common instances of each user are listed in Table IV (example: Among the proposals made by users 1 and 2 in Table III, IV instances are identical). Table V gives information on the results of the similarities between the dynamic profiles. To make recommendations of a semantic to a user group, only dynamic profiles that have a similarity greater or equal to the threshold are recommended for users concerned (see Figure 6), i.e., the profiles are considered similar according to the cosine similarity.

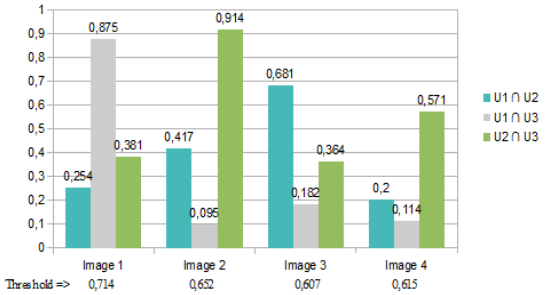


Figure 6: Similar dynamic profiles determination

The semantics of the image 1 can be recommended to users 1 and 3, the semantics of image 3 will be suggested to user 1 and 2, and those of image 2 to the users 2 and 3.

We compare our method with two classifications methods (SVM and Naive Bayes). The next table shows the result of our experiment. Three parameters are taken into account: the performance (possibility to reduce the errors) and classification time.

TABLE VI. COMPARING OF CLASSIFICATION APPROACHES

	Performance (Error Reduction)	(Classification time) (minutes))
Naive Bayes	89,72%	14,03
SVM	81,07%	09,62
Galois	91,18%	09,18

The experiment was conducted on 632 profiles of users on different classifiers. The results are presented in the Table VI. We note that all methods are efficient because they reduce significantly the error rate with different classification time. We find that the results of our approach are better because it can regroup, exhaustively objects in classes. Although there is a tradeoff between complexity and performance, it is still viable choices when better performance is considered.

VI. CONCLUSION

In this paper, we propose an efficient method for semantics recommendation based on indexing. We first formalize semantic indexing based on concepts of ontology. Then, we propose a similarity by exploiting the relationship between dynamic profiles. The similarities are used to make tighter recommendation of semantics to the users. The purpose of this recommendation system can be achieved through the management of static and dynamic profiles derived from semantic indexing.

The combination of semantic indexing and recommendation system calls for the development of more flexible recommendation methods that

 ISBN: 978-1-61208-305-6

allow the user to express the types of recommendations that are of interest to them rather than being “hard-wired” into the recommendation engines provided by most of the current vendors that, primarily, focus on recommending semantics to the user and vice versa. The second requirement of interactivity also calls for the development of tools allowing users to provide inputs into the recommendation process in an interactive and iterative manner, preferably via some well-defined user interface.

ACKNOWLEDGMENTS

This work has been funded by Electronic, Computer Science and Image Laboratory (LE2I) and by Doctoral School SPIM, FRANCE.

REFERENCES

[1] Fensel, D. et al, Spinning the Semantic Web:Bringing the World Wide Web to Its Full Potential. Cambridge, Mass.: MIT Press, eds, 2003.

[2] Haav, H. M., Lubi, T.-L.: A Survey of Concept-based Information Retrieval Tools on the Web. In Proc. of 5th East-European Conference ADBIS*2001, Vol 2., Vilnius "Technika", pp. 29-41.

[3] Guarino, N., Masolo, C., and Vetere, G. “OntoSeek : content-based access to the web*”.IEEE Intelligent Systems*, 14: pp. 70-80, 1999.

[4] K. Spärck Jones, IR lessons for AI, In Proceedings of Searching for Information, Artificial Intelligence and Information Retrieval Approaches, IEEE Special event, 1999. 245

[5] M. Mitra, C. Buckley, A. Singhal, C. Cardie, An analysis of Statistical and Syntactic Phrases, In Actes de la conférence Recherche d'Information Assistee par Ordinateur (RIAO), 1997.

[6] N. Hernandez, Ontologies de tâche et de domaine pour l’aide à l’exploration d’une collection de documents, In Actes du colloque de l'EDIT (Ecole Doctorale Informatique et Télécommunications), 2003.

[7] S. E. Robertson, K. Sparck Jones, Relevance weighting of search terms, Journal of the American Society for Information Sciences, 27 (3), pp. 129-146, 1976.

[8] S. Dumais, Latent Semantic Indexing (LSI) and TREC-2, In D. Harman (Ed.), The Second Text Retrieval Conference (TREC2), National Institute of Standards and Technology Special Publication 500-215, pp. 105-116, 1994.

[9] S. Dumais, Using LSI for information filtering: TREC-3 experiments, In D. Harman (Ed.), The Third Text Retrieval Conference (TREC3), National Institute of Standards and Technology Special Publication, 1995.

[10] D. Vallet, M. Fernández, P. Castells, An Ontology-Based Information Retrieval Model, In Proceedings of the 2nd European Semantic Web Conference, pp. 455-470, 2005.

[11] J. Kahan, M. Koivunen, E. Prud'Hommeaux, R. Swick, Annotea: An Open RDF Infrastructure for Shared Web Annotations, In Proceedings of the 10th International World Wide Web Conference, pp. 623-632, 2001.

[12] J.Paralic, I.Kostial, Ontology-based Information Retrieval, In Proceedings of the 14th International Conference on Information and Intelligent Systems, ISBN 953-6071-22-3, pp. 23-28, 2003.

[13] P. Zweigenbaum et al., Linguistic and medical knowledge bases: An access system for medical records using natural language, Technical report, MENELAS: deliverable 9, AIM Project A2023, 1993.

[14] M. Baziz, M. Boughanem, N. Aussenac-Gilles, C. Chrisment. Semantic Cores for Representing Documents in IR, In Proceedings of the 20th ACM Symposium on Applied.

[15] Ning RUAN, Ning HUANG, Wen HONG, Semantic-Based Image Retrieval in Remote Sensing Archive: An Ontology Approach, 2006.

[16] Haav, H. M., Lubi, T.-L.: A Survey of Concept-based Information Retrieval Tools on the Web. In Proc. of 5th East-European Conference ADBIS*2001, Vol 2., Vilnius "Technika", pp. 29-41.

[17] A. Kiryakov, B. Popov, I. Terziev, D. Manov, D. Ognyanoff, Semantic annotation, indexing, and retrieval, Journal of Web Semantics, 2(1), 2004.

[18] R.V. Guha, R. McCool, E. Miller, Semantic search, In Proceedings of the 12th International World Wide Web Conference, pp. 700-709, 2003.

[19] L. Khan, F. Luo, Ontology Construction for Information Selection, In Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence, pp. 122- 127, 2002.

[20] Woods, W., 97: Conceptual Indexing: A Better Way to Organize Knowledge. Technical report SMLI TR-97-61, Sun Microsystems Laboratories, Mountain view, CA.

[21] Armelle Brun, Ahmad Hamad, Olivier Buffet, Anne Boyer, Vers l’utilisation de relations de préférence pour le filtrage collaboratif, 17eme congrès francophone Reconnaissance des Formes et Intelligence Artificielle - RFIA 2010.

[22] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems : A survey of the state-of-the-art. IEEE Transactions on Knowledge and Data Engineering, 17(6) : pp. 734–749, 2005.

[23] M. Stolze, and M. Stroebel, “Dealing with Learning in eCommerce Product Navigation and Decision Support: The Teaching Salesman Problem,” 2nd World Congr. Mass Custom. Person., Munich, Germany, 2003.

[24] An-Te Nguyen, Nathalie Denos, Catherine Berrut, Bich-Thuy Dong Thi. Modèle de formation multiple de communautés dans un système de recommandation hybride.

[25] U. Shardanand and P. Maes. Social information filtering: algorithms for automating "word of mouth". In Proc. of the ACM CHI’95 - Conference on Human Factors in Computing Systems, volume 1, pp. 210–217, 1995.

[26] Michael J. Pazzani, Daniel Billsus, Content-Based Recommendation Systems, pp. 325-341, The Adaptive Web, Peter Brusilovsky, Alfred Kobsa, Wolfgang Nejdl (Ed.), Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, Lecture Notes in Computer Science, Vol. 4321, May 2007, 978-3-540-72078-2.

[27] A. Chandel, O. Hassanzadeh, N. Koudas, M. Sadoghi, and D. Srivastava. Benchmarking declarative approximate selection predicates. In SIGMOD'07, pp. 353-364, 2007.

[28] E. Spertus, M. Sahami, and O. Buyukkokten. Evaluating similarity measures: a large-scale study in the orkut social network. In KDD '05, pp. 678-684, 2005.

[29] Wille, R. (1980). Restructuring lattice theory, Ordered sets I. Rival.

 ISBN: 978-1-61208-305-6

[30] Wille, R. (1984). Line diagrams of hierarchical concept systems. Int. Classif. 11, pp. 77–86.

[31] Wolff, K. E. (1993). A first course in formal concept analysis - how to understand line diagrams. In F. Faulbaum (Ed.), SoftStatt'93, Advances in Statistical Software 4, pp. 429-438.

[32] Barbut, M. et B. Monjardet (1970). Ordre et classification, Algebre et combinatoire, Tome 2. Hachette.

[33] Birkhoff, G. (1940). Lattice Theory, Volume 25. New York : American Mathematical Society.

[34] J. Villerd, S. Ranwez, M. Crampes, Navigation sur des Cartes de Connaissances supportées par un Treillis de Galois Colloque Carto 2.0, Noisy-le-Grand, France, April 3, 2008.

 ISBN: 978-1-61208-305-6

Performance Analysis of the Opus Codec in VoIP Environment Using QoE Evaluation

Péter Orosz, Tamás Skopkó, Zoltán Nagy, and Tamás Lukovics
Faculty of Informatics
University of Debrecen
Debrecen, Hungary
e-mail: oroszp@unideb.hu

***Abstract*—VoIP has been a focus area of network communications for more than a decade now. The presence of VoIP traffic becomes more and more significant in the global Internet traffic. Although available access bandwidth is constantly increasing, higher capacity itself cannot guarantee higher quality of experience of the VoIP service. While QoE predicting methods are under active research, audio codecs also evolved greatly. The introduction and standardization of the Opus codec in 2012 is an important milestone in the voice codec evolution and Opus will probably be a royalty-free alternative for many VoIP applications in the near future. Past studies showed that audio quality of the Opus codec is superior when compared to almost every alternative. Mean Opinion Score (MOS) is a standardized scale for rating service quality. Our paper investigates the Opus codec in VoIP environment in terms of the relation between measured network QoS parameters and MOS value gained by subjective QoE assessments.**

Keywords—Opus codec; Speex codec; VoIP communication; speech quality; MOS; QoE evaluation.

I. INTRODUCTION

IP is a more and more preferred protocol for digital communication services and digital speech transport on IP (VoIP) is more and more significant in the global Internet traffic [1]. In the last decade, the evolution of real-time transport protocols and voice codecs resulted on an augmented user expectation in terms of service and speech quality. Ensuring high quality speech transmission is not a trivial task, since the service has to suit strict timing criteria. Voice communication is interactive and low latency (≤150 ms) is therefore a crucial requirement for the acceptable level of user experience. While mobile telecommunication companies operate dedicated infrastructure for speech transmission, provider independent VoIP sessions flow through heterogeneous public networks. It is more challenging to provide the sufficient level of service quality on a best effort infrastructure such as the Internet. Researches also point out that increasing bandwidth not necessarily ensures better subjective quality of the services.

Although the progression of VoIP technology is most visible on desktop platforms, an increasing number of users want to access these services using mobile devices. Lower computing performance and limited battery capacity make low code complexity a huge advantage for a voice codec. Simple PCM coding algorithms (like G.711 μLaw) were used for digital speech transmission from the beginnings. However, the increasing computing power in the last decade made possible to apply complex voice coding algorithms.

II. EVALUATING SPEECH QUALITY

Codecs tolerate network transmission anomalies (i.e., jitter or packet loss) differently and their behaviors have a direct impact on the subjective Quality of Experience (QoE). Methods for predicting QoE are under active research and development. In these methods, measured flow level QoS parameters (delay, jitter, reordering, packet loss) are associated with metrics based on subjective quality evaluation of service users. After call termination, providers often ask their users about the quality of the recent call. The most popular form for the evaluation is the 5-score Mean Opinion Score (MOS) scale [2]. Of course, this simple scoring scheme makes no relation between the quality of experience and the effect of different network anomalies. More detailed assessment techniques can provide better correlation but in practice, users cannot be asked for detailed and reliable evaluation easily. Other methods are predicting subjective quality experience estimation applying mathematical statistical evaluation on the received audio samples. The International Telecommunication Union (ITU) has its own recommendation for standardized evaluation of speech quality: after many years of development (superseding PEAQ, PSQM, PESQ and PESQ-WB algorithms), POLQA (ITU-T P.863) is able to evaluate speech sampled up to 14 kHz using the MOS metrics [3].

QoE prediction methods and algorithms are based on the statistics of a large measurement dataset. Some methods require the original audio data, these are called Full-Reference (FR) designs, while methods not requiring the original material are No-Reference (NR) type ones. In practice, acquiring the audio flow is often not possible or not applicable (lack of full control of endpoints, storage capacity limitations, privacy restrictions), and therefore, constructing a reliable NR method is consequential.

The introduction of the Opus audio codec is a significant milestone in world of voice codecs. The codec standardized by the IETF in 2012 is derived from the combination of the previously existing SILK (focusing on speech transmission) and CELT (aiming low latency) codecs [4]. Like most advanced codecs, it supports constant as well as variable

 ISBN: 978-1-61208-305-6

bitrates, and switching between rates with seamless transition. This feature makes possible to feedback altering network conditions (conjunction with Real-time Transport Protocol (RTP) [5] and RTP Control Protocol (RTCP) [6]). It is also effective for creating short audio clips because its algorithm does not need large code tables. Furthermore, it features advanced error correction. The correlation between audio frames can be adjusted, which controls how loss of audio frames affects voice quality. Also, optional Forward Error Correction (FEC) inserts redundant data (at the cost of some quality) to reduce the effect of packet loss. Opus is a new generation, universal audio codec, which is royalty-free in its every part. Its feature set, open source and industry support make presumably a popular audio codec for digital audio transmission over IP. In parallel with the standardization, a reference implementation of the codec library (i.e., encoder and decoder) is also developed and is freely available, which enables to evaluate the real-life performance of the Opus codec very effectively [7].

Although the audio quality was formerly evaluated (see Section III), the behavior of Opus codec under different network conditions has still to be investigated. We will sum up the works related to the Opus codec in Section III. In Section IV, a measurement setup for evaluating Opus in VoIP environment will be presented. The Opus codec had to perform on an emulated long distance network path implemented by our laboratory transport infrastructure. In Section V, a comparative performance analysis (set against its predecessor, the Speex codec) will be presented. Finally, Section VI concludes the presented work.

III. RELATED WORKS

Anssi Ramö and Henri Toukomaa evaluated Opus MDCT and LP modes with subjective listening tests and compared them with 3GPP AMR, AMR-WB and ITU-T G.718B, G.722.1C and G.719 codecs [8]. The paper keeps the codec a good alternative for the aforementioned codecs. The papers of C. Hoene et al. include different listening tests and compares the codec to Speex (both NB and WB), iLBC, G.722.1, G.722.1C, AMR-NB and AMR-WB [9][10][11][12]. They conclude that Opus performs better, though at lower rates, AMR-NB and AMR-WB still outperform the new codec. Jean-Marc Valin et al. present further improvements in the Opus encoder that help to minimize the impact of coding artifacts [13].

One of the motivational reason was that none of the researches above tested the Opus codec in VoIP environment. Moreover, the original or input audio signal is usually not available. Therefore, currently available FR-based methods cannot be applied by service providers. Our aim is to construct a NR method for predicting subjective QoE based upon measured QoS parameters.

IV. MEASUREMENT SETUP

An emulated long distance network path including two communication endpoints was constructed for the assessment of both codecs (Fig. 1). Endpoints feature generic multi-core x64-based architectures. They were equipped with Intel PRO/1000 NICs and interconnected with 2 m of industrial grade CAT6 cabling. Fedora Core 18 was installed to both hosts with unmodified Linux 3.8.1-x kernel (with a jiffy setting of 1000 Hz). We have chosen version 1.2.2 of the sflPhone VoIP application, since it supports Speex as well as Opus and its transmission parameters conformed the expected QoS performance (packet rate, uniform distribution of inter-arrival times and packet sizes) [14].

A carefully selected audio clip (easy to understand single channel of speech) was injected into the input of the softphone on Host A. JACK Audio Connection Kit is a general audio tool and is able to connect audio inputs and outputs of different applications and audio devices [15]. Current version of sflPhone can accept ALSA and PulseAudio datastream at its input. PulseAudio was selected since it can be directly connected with JACK. Since it has native output plugin (sink) for JACK, the audio clip was fed into JACK from an uncompressed PCM WAVE file with the GStreamer application. We carefully configured the applications not to perform unnecessary audio sample rate conversion throughout the digital audio path. The sflPhone application on Host B was configured to save the audio data into uncompressed PCM WAVE file for further QoE assessment. During the measurement we used the netem Linux kernel module, which was configured symmetrically on both directly connected interfaces to emulate a long distance path and produce various network anomalies that affect QoS (i.e., packet loss and variation of delay (jitter)). During the measurements we stored both the WAVE file from the receiver softphone and the PCAP files containing the received RTP stream [16]. The first 35 seconds of the original speech were used as input in all measurements. The network delay was set to 100 ms in each direction. The codecs were measured independently from each other, with the same series of parameters (Table 1). Netem network parameters were iterated using the following scheme:

TABLE I. MEASUREMENT PARAMETERS

Measurement series	**Opus**	**Speex**
Jitter (ms)	0, 1, 2, 3, ..., 20	
Packet loss (%)	1, 2, 3, ..., 40	
Combined: jitter (ms) and packet loss (%)	jitter: 1, 2, 3, ..., 10 loss: 1, 2, 3, …, 10	

The measurement sessions resulted in 160 audio clips per codec. As a reference of the evaluation, an initial measurement with zero jitter and no packet loss were run for both codecs.

 ISBN: 978-1-61208-305-6

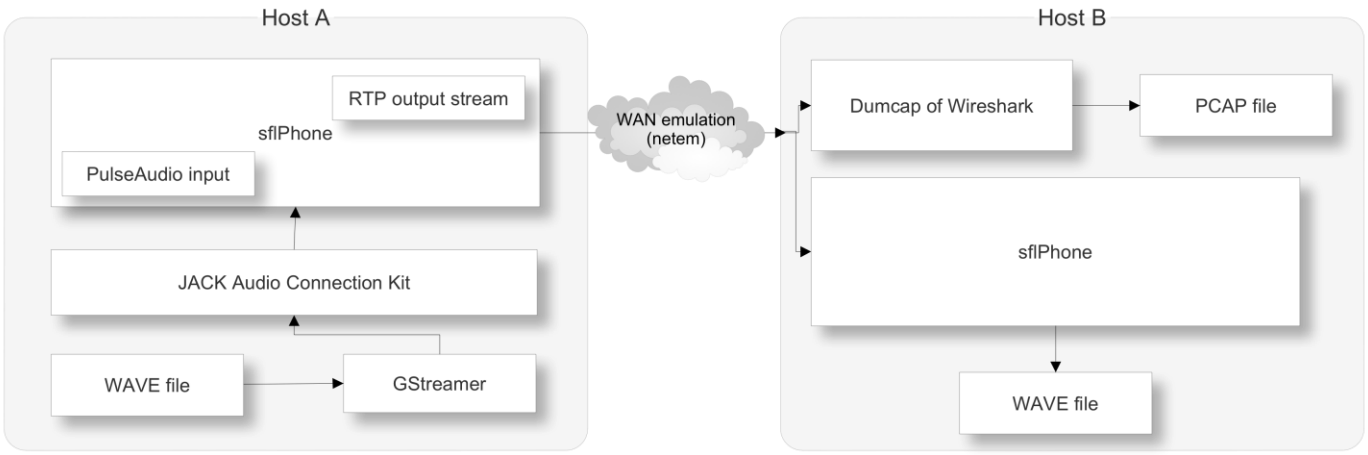


Figure 1. The measurement setup: audio is fed into the VoIP client on Host A and is transported through RTP to the other client on Host B.

V. EVALUATION OF THE MEASUREMENTS

We used 16 kHz sample rate for both codecs, since wideband (WB) operation mode is now a reasonable user claim. Both codecs were operated in variable bitrate (VBR) WB mode during the measurements. In case of the Opus testing, sflPhone generated 100 RTP packets per second, with variable packet size from 40 to 159 bytes that are sent out with a 8 ms (with a standard deviation of 500 µs) period. Opus was set to constraint VBR mode when the encoder assumes a transport with an average of the nominal bitrate and it creates one frame for the corresponding buffering delay (Fig. 2). The nominal bitrate was 64 kbps during the Opus measurements.

In case of Speex, 50 RTP packets were sent out per second, at an average period of 18 ms (with a standard deviation of 1 ms) and packet size was fixed to 124 bytes. The average bandwidth was 42 kbps (Fig. 3). Speex calls this setting "wideband". With a typical consumer access bandwidth, it is reasonable using such or even higher quality settings.

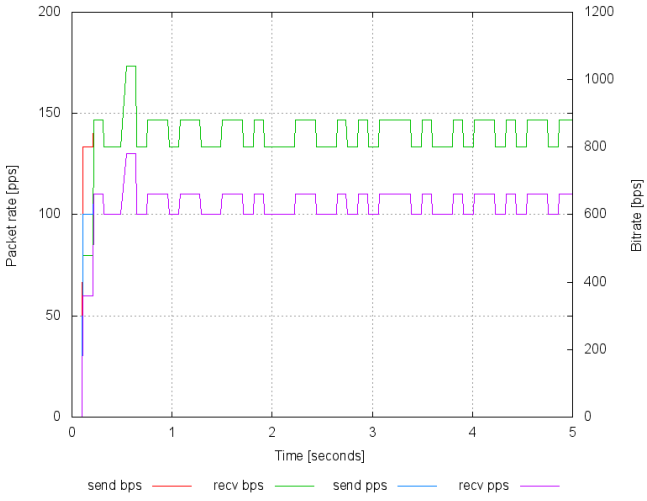


Figure 2. Opus: Packet rate and bandwidth during the voice transfer

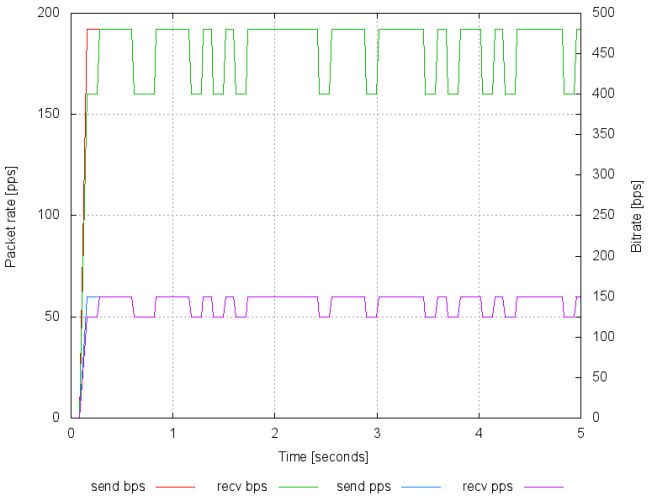


Figure 3. Speex: Packet rate and bandwidth during the voice transfer

All of the captured audio files were sequentially listened by users with sufficient amount of time for relax. The files were graded using the 5-point MOS scale.

A. *Jitter-sensitivity*

Under normal conditions, packets should arrive in a restricted time window to the decoder to maintain the real-time service. Variance of packet arrival are caused by infrastructural delay (routers can have queues with different priorities for forwarding) or by transient (longer burst than internal buffers allow) overload of the receiving endpoint. The jitter buffer of a real-time application is for holding the incoming packets and eliminating network jitter introduced by the infrastructure. The size of this buffer should be kept relatively small for the VoIP applications to achieve the required low latency performance. Packets arriving out of the expected time range are dropped.

Opus codec seems to be more sensitive to jitter but performs better than Speex at extreme conditions (see Fig. 4). Opus produced better voice quality at low jitter. Furthermore, even at 11 ms of jitter, the decoded voice was still more understandable than with Speex. None of the users gave 5 points for the Speex performance even at the smallest amount of jitter (1 ms) since it caused not annoying but clearly audible clicks. From the aspect of jitter, Speex gives average

 ISBN: 978-1-61208-305-6

performance at a wider range but Opus provides higher voice quality under 4 ms of jitter and is easier to listen to under heavier network perturbation.

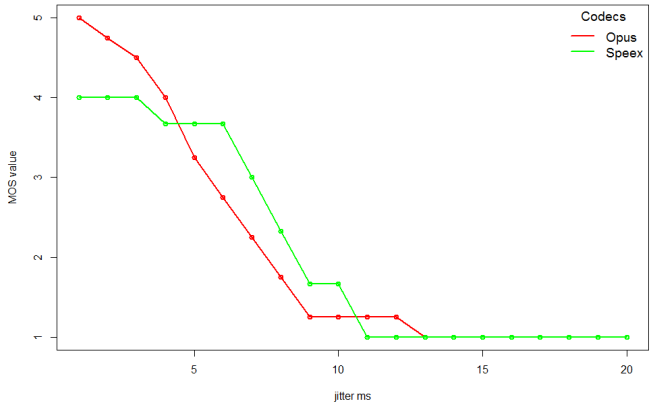


Figure 4. Correlation between jitter and subjective quality of experience expressed in MOS

B. *Loss-sensitivity*

Packets can be lost throughout the network path (e.g., inside a router) or at the endpoint itself. It is difficult to evaluate how efficient the codecs are in the compensation of information loss.

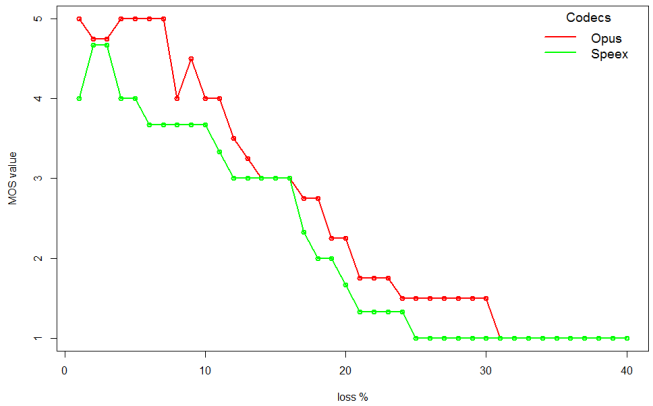


Figure 5. Correlation between packet loss ratio and subjective quality of experience expressed in MOS

As seen in Fig. 5, the Opus codec smoothes the effect of packet loss more efficiently. While Speex is gibberish even at 25% packet loss (using 802.11 access, it is not an unrealistic situation), Opus still gives acceptable result up to 30% of loss. Further observation is related to the split opinions at low packet loss: Opus codec performed better at 2% of loss that at 1%. This result may reflect the fact that a higher performance psychoacoustic model is working inside the codec. At a particular loss, quality of experience with Opus decreases less than with Speex.

C. *Sensitivity for multiple anomaly*

Anomalies detailed in the previous subsections are rare to occur alone. In reality, some combination of jitter and packet loss should be expected. Accordingly, we executed a complex measurement session to evaluate the audible effect of the presence of both jitter and packet loss.

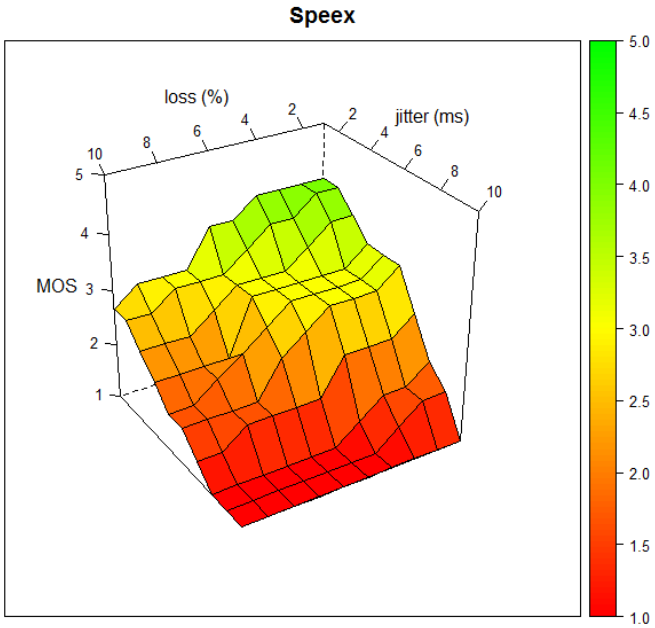


Figure 6. MOS values for the Speex codec under mixed network conditions

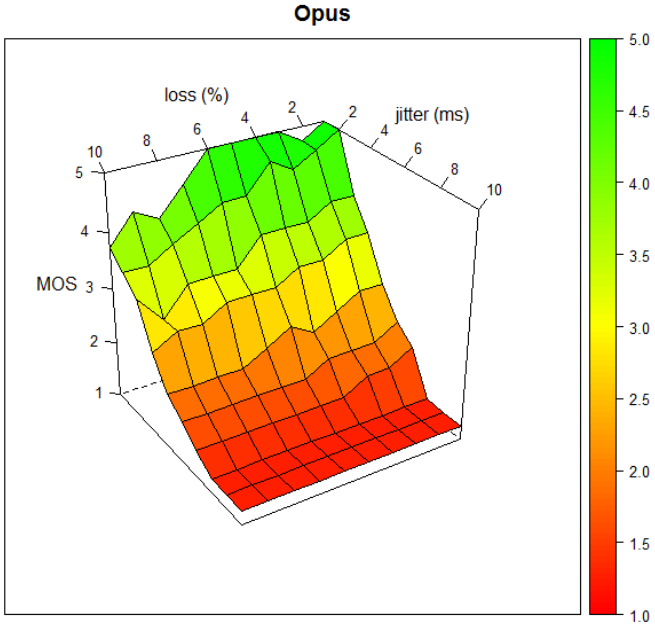


Figure 7. MOS values for the Opus codec under mixed network conditions

 ISBN: 978-1-61208-305-6

Figs. 6 and 8 show that Speex got 3 MOS points in a wide range of the investigated network parameters. In the MOS scale 3 points equivalent with the lower bound of the acceptable quality. It never reached 5 score even at small amount of anomaly. In contrast, Opus performs uniformly until its boundaries (see Fig. 7). Although jitter error affects its quality more drastically, it is more tolerant to loss than Speex. The QoS-QoE relationship of the Opus codec in terms of jitter is more close to linear than that of Speex (Fig. 7).

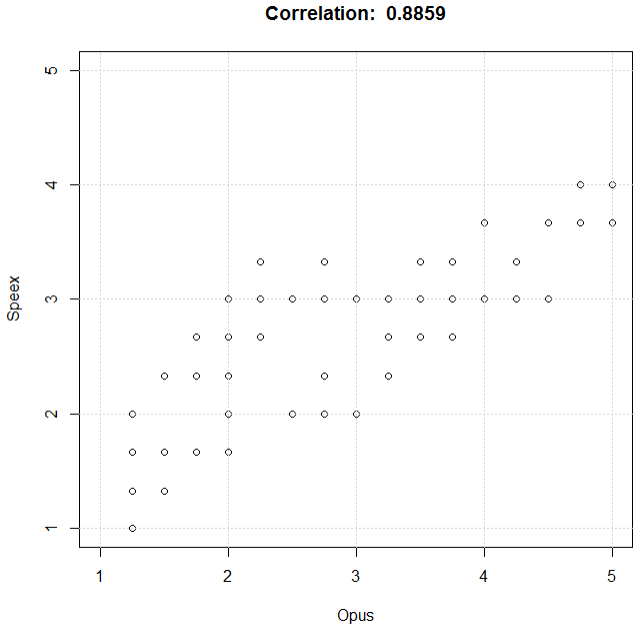


Figure 8. Correlation between Speex and Opus MOS scores for all of the combined measurement scenarios.

As presented earlier in this section, the QoE assessment assigned a MOS value for each measurement. According to our combined measurement series (both jitter and loss are present on the emulated network path) now we got 100 MOS values for both codecs. The correlation of the two MOS series, which is calculated from (1) is presented on Fig. 8.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_Y)]}{\sigma_x \sigma_y} \qquad (1)$$

where μ is the expected value of the random variable and a σ is its standard deviation. Using two variable polynomial regression and Sum of Squares due to Error (SSE) goodness-of-fit statistics, we found that jitter and MOS values show a linear relation, while loss and MOS values suggest a quadratic relationship. In the near future, our goal is to construct a low order estimator function for calculating MOS values based on packet level QoS parameters (i.e., loss and jitter). This estimator function could be the basis of a NR-type objective QoE assessment method for Opus based VoIP conversations.

VI. CONCLUSION

In this paper, the fault tolerance of the royalty-free Opus and Speex VoIP codecs has been evaluated using laboratory QoS measurements and subjective QoE assessments. Although their roots are the same, under the investigated conditions Opus performs more uniform when multiple network anomalies of jitter and packet loss are present. Since there is no NR method available for speech quality prediction available, close-to-linear relationship between measured jitter and the gained subjective QoE values of Opus codec make possible to create a NR method to estimate QoE from the measured QoS parameters. We are actually moving this way on. We also note that Opus' Forward Error Correction option for transmitting redundant information is another important feature that has to be evaluated in a future work.

ACKNOWLEDGMENT

The work was supported by the TÁMOP 4.2.2.C-11/1/KONV-2012-0001 project. The project was implemented through the New Széchenyi Plan, co-financed by the European Social Fund.

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4.A/2-11-1-2012-0001 'National Excellence Program'.

REFERENCES

[1] Point Topic Ltd.,VoIP Statistics – Market Analysis, Q2 2012, October 2012

[2] ITU-T P.800: Methods for subjective determination of transmission quality, August 1996

[3] ITU-T P.863: Perceptual objective listening quality assessment, January 2011

[4] JM. Valin, K. Vos, and T. Terriberry, "IETF RFC 6716: Definition of the Opus Audio Codec", September 2012

[5] Real-time Transport Protocol, http://tools.ietf.org/html/rfc3550 [retrieved: September 2013]

[6] RTP Control Protocol, http://tools.ietf.org/html/rfc3550 [retrieved: September 2013]

[7] Opus Codec Downloads, http://www.opus-codec.org/downloads/, [retrieved: August, 2013]

[8] A. Ramö and H., "Voice Quality Characterization of IETF Opus Codec"INTERSPEECH, pp. 2541-2544. ISCA, 2011

[9] JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-00" June, 2011

[10] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-01", May, 2012

[11] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-02", May, 2012

[12] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-03", November, 2013

[13] JM Valin, G Maxwell, TB Terriberry, and K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec, " 135th AES Convention , October 2013

[14] sflPhone, http://sflphone.org/, [retrieved: August, 2013]

[15] JACK Audio Connaction Kit, http://jackaudio.org/, [retrieved: August, 2013]

[16] J. Spittka, K. Vos, and JM. Valin, "RTP Payload for Opus Speech and Audio Codec draft-ietf-payload-rtp-opus-01", August 2013

 ISBN: 978-1-61208-305-6

Multi-threaded Packet Timestamping for End-to-End QoS Evaluation

Péter Orosz and Tamás Skopkó
Faculty of Informatics
University of Debrecen
Debrecen, Hungary
e-mail: {oroszp, skopkot}@unideb.hu

Abstract— **In this work, we are focusing on the enhancement of end-to-end QoS evaluation by improving the performance and the functionality of packet timestamping. Accordingly, a new software-based multi-layer timestamping method is introduced, which implements a multi-threaded offloading mechanism. Compared to the available generic kernel-time based solutions, it provides not only higher precision but also lower kernel level overhead and thus lower packet loss without using any special hardware component. These improvements make the proposed method a more efficient basis for multi-layer QoS measurement performed on-the-fly on the communication endpoint, which may result in a better QoS-QoE correlation. The efficiency of the solution is validated against the generic, kernel-time based timestamping using their Linux implementations.**

Keywords-Timestamping; Media traffic; QoS evaluation; Next Generation Networking; Computer network management; Network measurement; Linux

I. INTRODUCTION

With the emergence of 1 Gbit/s and faster access networks and with the increasing demand for their packet level monitoring, the dominance of the purely software based network measurement tools, operating on generic desktop PCs, gradually decreased. Considering the high transmission rate on the monitored network link, the resolution and precision of software timestamping methods and the lossy packet procession provided by a generic Network Interface Card (NIC) became a serious bottleneck of traffic analysis. The drawbacks of the software-based packet capturing are already investigated by several papers (see Section II). As primary effect, using low resolution and precision software timestamps leads to an incorrect representation of the packet inter-arrival times, since the generation of these timestamps is performed within a shared resource environment, where the timestamping process, as any other process, is scheduled by the OS scheduler subsystem and competes for CPU time. The common kernel-time based solutions operate with large overhead and high variance, which could be a bottleneck of precise QoS measurement. The final 64-bit Time of Day (ToD) format of the software timestamp is calculated within the kernel space on-the-fly, based upon an arbitrary kernel clock-source (High Precision Event Timer (HPET), Advanced Configuration and Power Interface (ACPI), Timestamp Counter (TSC), etc.), which, by executing several conversion functions upon packet reception, results in a large packet processing overhead. Among other complex packet processing tasks, the timestamp calculation is implemented within the packet reception softIRQ, which, due to its complexity, produces excessive CPU load. The OS scheduling mechanism and the large timestamping overhead provides a very low precision timestamping output, while the intensive CPU usage of softIRQ results in packet loss. All these effects make end-to-end Quality of Service (QoS) evaluation that includes flow level delay, jitter, packet loss, and reordering measurement, very ineffective on high speed communication links. Nevertheless, today's hardware accelerated network monitoring devices are designed to operate on aggregated backbone link, not on an access link belonging to one single endpoint node. In contrast, monitoring QoS level of real-time media services on the communication endpoint itself should be a reasonable option and this is the point where the software-based on-the-fly QoS evaluation comes in. However, QoS evaluation (including timestamping) requires CPU and other resources and therefore it should be performed with low overhead without degrading the performance of the monitored real-time media communication.

We are facing three rudimental problems: low resolution and low precision of the timestamps, and large overhead. The resolution of the software timestamp, as in case of any timestamping mechanism, depends on the resolution of the applied hardware clock source, the length of the data structures that store the generated timestamp value, and the granularity of the clock-to-time conversion. Enhancing the microsecond resolution up to one nanosecond is not a particularly big challenge, since today's x86 and x64 CPUs operate at 1+ GHz and support the constant Time Stamp Counter (TSC) register, which acting as a kernel clock source enables the timestamping subsystem to generate timestamps with 1 ns resolution [1]. The constant rate property assures the register value to be incremented with a fixed rate according to the maximum CPU frequency, independently of the current operational mode.

High resolution time measurement is already supported by the Linux kernel from version 2.6.27 [2]. All we have to do is to provide an unconverted 64-bit path for these high resolution timestamps up to the user space monitoring application. To achieve this goal, we previously enhanced the common libpcap library to natively handle the nanosecond resolution timestamps provided by the kernel [3][4][5]. Unfortunately, the enhancement of the resolution alone does not imply high time precision. The primary bottleneck - the large deviation of the generation overhead - decreases the precision to an unacceptable level. Moreover,

 ISBN: 978-1-61208-305-6

the large CPU overhead of the built-in timestamp generator mechanism could lead to a serious amount of packet loss during the real-time conversation. To overcome this problem, a new timestamping method was designed and implemented, which is based on a multi-threaded offloading approach. The timestamping process is split into two separate phases. The primary goal was to achieve minimal timestamping overhead in kernel context with very low variation (see Section III). Even if we know that it is impossible to provide a precision close to the hardware-based timestamping solutions, a realistic goal was to significantly exceed the precision level of the existing software-based solutions available on generic multi-core architectures. A hardware accelerated approach of multi-layer timestamping is already presented in our previous paper [6]. However, in this proposal, we introduce a purely software based timestamping solution with low overhead and low variance, which enables to apply the method for multi-layer timestamping on generic PCs without any hardware acceleration.

Another critical aspect of packet processing is the packet loss ratio caused by the capturing itself. Beyond the benefits related to timestamp precision, the proposed timestamping method significantly decreases the packet loss compared to the large overhead kernel-time (ktime) based timestamping mechanisms. Active QoS monitoring of real-time media applications can therefore benefit from this method.

The rest of the paper is organized as follows. Section II gives a summary of related works in the field of high precision software-based packet timestamping. Theoretical background of our multi-threaded timestamping method is described in Section III and its Linux based implementation is presented in Section IV. We evaluated the performance of the introduced method using comparative laboratory measurements in Section V. Finally, Section VI concludes the presented work.

II. RELATED WORKS

In the last decade, several research projects focused on the challenges of high performance network monitoring, especially triggered by the emergence of the 1+ Gbps networking technologies [7][8]. While most of them are hardware accelerated solutions, some projects investigated the possible performance enhancement of the software-based network monitoring suites. Coppens et al. introduced a new scalable network monitoring platform called SCAMPI [9], which supports dedicated hardware accelerated monitoring boards as well as generic NICs for packet capturing. Heyde et al. investigated the loss property of the Intel NIC-based packet capturing in the context of Lawful interception [10]. Pásztor et al. presented a high resolution, low overhead timestamping method based on the CPU's TSC register [11]. Their timestamping proposal includes an offloading mechanism based on a post-processing phase. In our work, we defined two parallel goals: improving the performance of the offloading method proposed by Pásztor et al., and also decreasing the packet loss ratio during the capture process. The TSC clock source has high resolution as well as high precision, as already investigated in [11]. However, the software timestamping methods, relying on the TSC register as clock source, have a very limited overall precision due to the large generation overhead and the OS scheduling (within the shared resource environment). These bottlenecks do not enable these methods to provide adequate precision for high speed QoS evaluation.

III. THEORETICAL BACKGROUND

Our proposed timestamping mechanism's primary benefit is its ultra-low timestamp generation overhead and the low variance of this overhead on most of the generic purpose multi-core system. The applied clock source is the TSC clock cycle register, which is read by a custom, high priority kernel process at packet arrival, then, the obtained value is passed to the higher level packet processing application, which offloads the clock-to-time conversion. The original method proposed by Pásztor et al. performed the conversion in a separated post-processing phase and did focus neither on enhancing the packet processing performance nor its multi-layer application. In contrast, our overhead reduction and precision improvement is gained by the combination of following two ideas.

A. Decreasing timestamping overhead and packet loss ratio within the kernel space

The packet processing softIRQ, which implements kernel-level timestamping functionality, should be statically assigned to an otherwise idle CPU core by directly altering its core affinity. Instead of providing the final timestamp format, timestamping within the softIRQ context should include the acquisition of the TSC value only, which requires not more than 24 clock cycles to be performed. Then, this 64-bit clock cycle value, which represents the moment when a packet reaches the kernel's network stack, is preserved with the packet within its path up to the conversion thread. For further improvement of timestamp precision, all interrupts are disabled on the assigned CPU core during the execution of the timestamping code.

B. Offloading timestamp conversion to the user space

The cycle-to-time conversion should be done on-the-fly by a dedicated user space thread of the multi-thread capturing process. After conversion, the nanosecond resolution timestamps, which is now in time of day format, should be sent back to the main packet processing thread. The dedicated timestamp conversion thread is also running on an otherwise idle CPU core and therefore it provides a very low conversion overhead (Fig. 1).

The combination of these two ideas does not just reduce timestamping overhead and improve the precision, but significantly decreases the kernel-level packet loss ratio at high arrival rates: by offloading the cycle-to-time conversion, the low level processing of each incoming packet requires lower CPU resource and thus, more packets can be accepted by the kernel networking subsystem within the same time interval without resource exhaustion. To maintain the low loss ratio up to the capture application, large packet buffers should be applied at the higher levels of the kernel space. Our comparative measurements validate the

 ISBN: 978-1-61208-305-6

performance parameters of the proposed method using its Linux based implementation (see Section V).

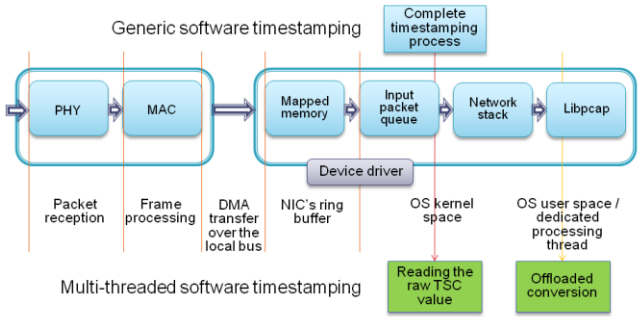


Figure 1. The new timestamping method produces a raw timestamp value in kernel space with very low overhead and implements the offloading of the cycle-to-time conversion by a dedicated processing thread in user space.

In network measurement, the precision of timestamping is a criterion more important than the low clock offset, especially for measuring packet inter-arrival times and round-trip delays at one single point of the network (e.g., active probing). On a generic purpose PC without any additional hardware, the TSC clock cycle register is the optimal choice for high precision packet timestamping, since it has a very low rate error (< 0.3 ppm) as presented in [11]. The multi-core architecture and the new multi-threaded offloading method together can provide enough processing capacity to fulfill the requirements of high precision packet timestamping.

The proposal of Pásztor et al. is based upon Fast Ethernet measurements. Today, Gigabit Ethernet is widely used for connecting endpoints to the network. Our multi-threaded timestampig solution was therefore validated using intensive traffic patterns on 1 Gbps network links. Intensive traffic on a Gigabit Ethernet link may involve high packet rate with inter-arrival times below 1 μs. The timestamping mechanism must provide enough resolution and precision to realistically interpret packet arrivals even at high rate. The resolution and precision capabilities of the software-based timestamping are determined by the generation overhead and its variance. Accordingly, the desirable overhead of packet timestamping should be well below the 1 μs order to represent the inter-arrival times.

On a single core system, even with our multi-threaded offloading method, it is hard to demonstrate such a low timestamp generation time since it is not possible for the packet processing softIRQ to be assigned to an idle CPU core. Accordingly, we assume, that a generic multi-core CPU with constant TSC register is available for the measurement task.

IV. Implementation of the Proposed Method

We decided to implement our method under Linux since it has sophisticated interrupt handling and scheduling mechanisms. As a first step, we modified the kernel's packet timestamping code by replacing the complex and relatively time-consuming calls by a simple RDTSC instruction. In the unmodified kernel, timestamps are generated by calling ktime_get_real() that queries the system's clock source and calculates a wall-clock time. RDTSC is an x86 CPU instruction that reads the CPU's TSC register. When compared to the built-in timestamping method, executing RDTSC takes shorter time, just about 24 clock cycles on a decent Intel CPU. Since the generated values are not represented in wall-clock format, they need to be post-processed later.

As a further optimization, we locked the packet processing softIRQ to a specific CPU-core. It ensures that no other interrups are intefering with timestamping and the packet processing functions. Multiple CPU cores don't necessarily keep their TSC counters in sync. Locking the softIRQ to a specific core also ensures that the read TSC values are acquired from only one CPU. The cycle to ToD conversion should be done in the user-space, utilizing another core. The off-line processing is elaborate when doing long-term capturing. Therefore, we modified the libpcap library to do the conversion at packet receiption. The capture application should query the CPU frequency and report it to libpcap to be able to the conversion. We locked the dumpcap application to another core. By using a multi-threaded design, libpcap could utilize more than one core for converting the timestamps.

V. Evaluation of the Timestamping Performance

Comparative measurements and statistical analysis of the measurement output data are used to validate the performance parameters of the new multi-threaded offloading method for software timestamping. The investigated parameters for both methods (the kernel-time based and the multi-threaded one) involve the per-packet overhead of the timestamping process, the variance of the overhead for the processed packets and the overall packet loss ratio within the system. A generic purpose PC with Intel Core i7 K-2600 2.93 GHz CPU and Linux non-preemptive kernel 2.6.39.2 was set up for all of our measurements. The used NIC was the common and well documented Intel 1000/PRO PT PCI Express card with the e1000e Linux device driver version 1.3.10-k2. This driver implements the New API (NAPI) operation mode introduced by the Linux kernel from version 2.4, which determines the low-level packet processing mechanism of the system [12]. All of the network stack related kernel buffer and driver parameters were optimized for intensive incoming traffic. With the following comparative measurement session, the characteristics of the timestamping overhead and its variation were investigated. In order to accurately measure the overhead of the timestamping mechanism, we modified the original Linux kernel. Two extra timestamping checkpoints had been inserted into the packet processing path, one just before the execution of the packet timestamping function __net_timestamp(), and another straight after its return point. These two timestamps are generated by the rdtscll() function call, which has a very low overhead, since it does not perform any conversion. The delta value minus the checkpoint generation overhead defines the per-packet timestamp generation overhead. This delta value is converted

 ISBN: 978-1-61208-305-6

to a real time value in a later evaluation phase. The built-in (kernel-time based) timestamping code performs real time cycle-to-time conversion and therefore, it produces an overhead with high variance, which results in low precision and also implies significant packet loss. Offloading this conversion task reduces timestamp generation time, which improves precision and reduces packet loss as well.

A. *Investigating precision: line-rate homogenous traffic pattern*

The first measurement scenario enables us to investigate the precision property of both timestamping methods. In this setup, we applied homogenous traffic patterns including fixed packet size and fixed inter-frame gap (IFG). This measurement type requires a high precision traffic generation device, which produces the line-rate packet stream in hardware. For this purpose, we applied a dedicated FPGA packet generator with Gigabit Ethernet interface. Since the generator and the measurement PC are directly connected, the packet inter-arrivals as seen on wire level are determined by the transmission rate of this device and therefore, it can be considered constant. The measurement PC, which was a general purpose desktop computer, performed software based packet capturing with a modified libpcap library supporting both timestamping methods: the generic ktime-based and the multi-threaded one. In both cases, the timestamping code is executed at a well defined point of the packet processing path. For non-NAPI supported device drivers, the software timestamping code is executed within interrupt context by the top half interrupt handler, when the packet is en-queued into the input packet queue of the operating system, while with NAPI support (it is our case), the timestamping is performed by the bottom half handler (softIRQ) [14]. In this latter configuration, the arrival time represented by a software timestamp indicates the moment when the packet is de-queued from the input packet queue, and therefore is affected by the scheduling mechanism of the Linux kernel and the intensity of the interrupts per second on the CPU core that the softIRQ is running on. According to the new method, interrupt handling was disabled during the execution of the kernel level timestamping code with the local_irq_disable() kernel function. This involves that no interrupt is generated for the affected CPU core until the interrupts are re-enabled. Since the execution overhead of the ktime-based method is higher than the multi-threaded one, the probability of an interrupt event (executing a top half handler) to happen during its execution is also higher. The execution of the top half handler puts the CPU in interrupt context, which causes jitter in the execution of the timestamping process.

The measurement (see Fig. 2 and Fig. 3) is performed with an artificially generated traffic pattern (pattern #1) that complies with one direction of a single VoIP conversation including 140 bytes packets transmitted by the traffic generator device in each 20 ms period, which is equivalent to 2,499,860 bytes of IFG. This is a typical packet size and transmission intensity provided by a VoIP audio codec, i.e., G.729.

The second measurement is also done by a synthesized traffic pattern of a HD video stream (see Fig. 4 and Fig. 5). Pattern #2 includes 1,368 bytes packets with a fixed 169,631 bytes of IFG value.

Since these traffics contain homogenous packet sequences, they are suitable to measure the timestamping methods in terms of the variance of the measured inter-arrivals, and the order and the variance of the generation overheads.

The synthesized VoIP traffic, due to its light packet-arrival intensity of 20 ms, does not imply high system load. However, during this measurement, the timestamping overhead provided by the generic ktime-based timestamping presents an average of 195 ns with a large deviation. In contrast, the multi-threaded timestamping method, by applying an effective timestamp acquisition method and offloading several conversion tasks to the user space, results in a low overhead, the half of that of the ktime-based one (Fig. 2).

TABLE I. TIMESTAMPING OVERHEAD FOR PATTERN #1

Timestamping method[a]	Overhead average [ns]	Overhead variance [ns]
Generic ktime	195.2762	176.0954
Multi-thread TS	73.49326	3.793385

a. Homogenous traffic of 140-byte packets with 2,499,860 bytes of IFG

Moreover, the variance provided by the overhead values is also significantly lower (Table 1 and 2). Considering the measurement results, we can find out that both the effective resolution and the precision of the timestamps are significantly improved with the new timestamping method. Fig. 2 represents some extreme high values in the inter-arrival times measured by the ktime-based method (green bars), while these large values are not present in the result of the multi-threaded measurement. The smaller deviation of the multi-threaded solution implies higher timestamping precision.

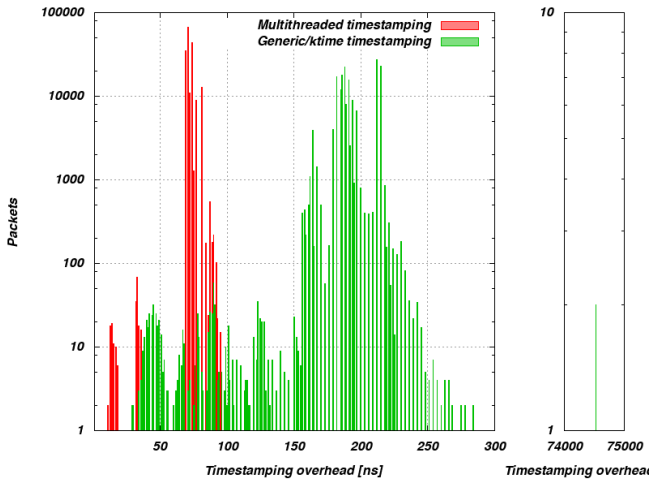


Figure 2. Density points of the timestamp generation overheads for the generic method and the multi-threaded one.

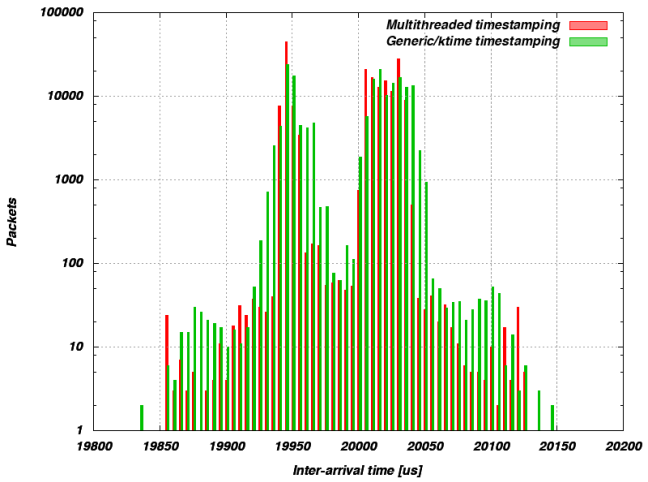


Figure 3. Density points of the packet inter-arrival times

We observed that the large inter-arrival values are in correlation with the large timestamping overhead values. Accordingly, we can assume that the large variance of the generation overhead drives to a very low precision representation of the packet inter-arrival times and could lead to false measurement results. Since there is a high correlation between the generation overhead and the measured arrival time, the low variance of the overhead is the key factor to get high timestamping precision. Whereas the overhead values of the generated traffic fall in a relatively small range, small histogram bins should be chosen.

This applies to the inter-arrival time graph, too. The histogram bars are spaced loosely and so, to make both measurement results represent on the same histogram.

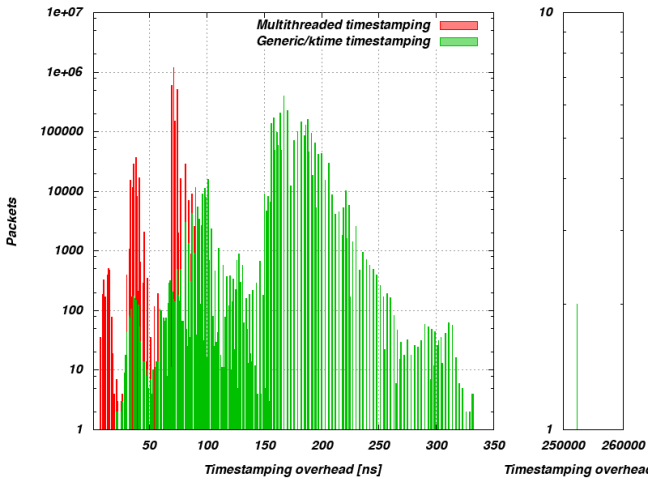


Figure 4. Density points of the timestamp generation overheads for the generic method and the multithreaded one.

Later in this section, we will investigate the root cause of the large variance represented by the generic ktime-based method.

TABLE II. TIMESTAMPING OVERHEAD FOR PATTERN #2

Timestamping method[a]	Overhead average [ns]	Overhead variance [ns]
Generic ktime	173.3606	177.1197
Multithread TS	70.84912	8.197374

a. Homogenous traffic of 1368-byte packets with 169,631 bytes of IFG

Nevertheless, the new multi-threaded timestamping method introduced in this paper, is a dedicated software-based method, that improves the resolution as well as the precision of the traffic measurement. Besides its primary benefits, its low generation overhead has also a positive side-effect: since it is less CPU intensive than the ktime-based one, capturing intensive network traffic will result in lower packet loss rate on any generic PC (see Fig. 6a and Fig. 6b). The auto-correlation analysis of the overhead series showed that the system with the new method has a high memory. Accordingly, based on the current overhead value, future overhead values can be predicted with a higher probability, which implies higher system stability and improved timestamp precision. If the intensity of the packet arrivals is higher, as with the second measurement, the difference in the measured properties becomes more obvious (see Fig. 4 and Fig. 5).

B. Investigating packet loss ratio within the system

We also investigated the packet loss rate during the capturing process in case of both methods. Extreme low overhead values (< 25 ns) derive from the instruction caching mechanism of the applied CPU architecture (Fig. 4). Each measurement round was performed with fixed packet size and fixed IFG combination. 20,000 packets were transmitted each time by the hardware accelerated traffic generator presented earlier, which was directly connected to the measurement PC. The generated packets were captured by the dumpcap application. According to the results presented in Fig. 6a and Fig. 6b, the loss ratio with the multi-threaded timestamping method is improved. Since this method uses a low overhead access to the TSC register and offloads all of the conversion tasks from the kernel space to a dedicated user thread.

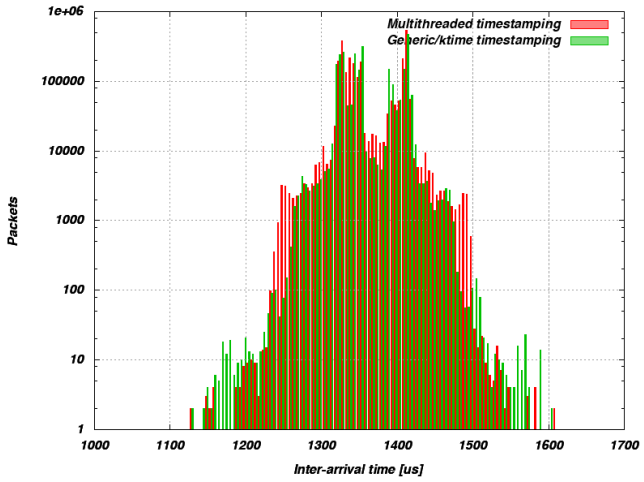


Figure 5. Density points of the packet inter-arrival times

This free time becomes available (within the kernel space) to the kernel scheduler to give it away to another CPU intensive softIRQ tasks that are parts of the packet processing subsystem.

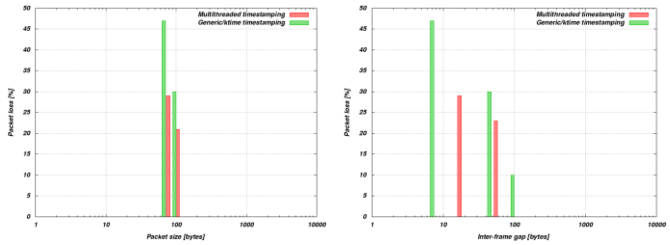


Figure 6. a) Packet loss ratio in function of the packet size b) Packet loss ratio in function of the inter-frame gap

Moreover, the execution of the conversion code in user space is more tolerant to delay. These benefits directly affect the packet processing performance and thus the loss ratio. When the CPU core, dedicated for the execution of the packet processing softIRQ, gets full load for a relatively long time period then the NIC's private queue reaches its maximum length and packet loss will happen.

VI. CONCLUSION

A multi-threaded timestamping method for end-to-end QoS evaluation has been introduced, which provides low kernel context overhead by eliminating the built in clock source API inside the kernel and offloading the conversion tasks to a dedicated processing thread in the user space. In contrast to the solution of Pásztor et al., this method implements a multi-threaded design that supports on-the-fly timestamp conversion and multi-layer application. The main goal of the design was to minimize kernel context timestamping overhead during packet capturing in order to improve timestamping precision and increase effective resolution. The performance properties of the new solution have been evaluated and compared against the generic kernel time based timestamping mechanism available in the Linux kernel. We showed that the overhead of our new method is half of the generic one enabling multi-layer timestamping and the variation of the overhead that determines the precision property of the timestamping is also significantly improved. Over the main benefits, this solution has a positive side effect: the packet loss rate, while capturing high intensity network traffic, is decreased, since the new method requires less CPU resource in kernel context, furthermore the execution of the offloaded timestamp conversion tasks can be delayed due to the scheduling policy of the user space. Though our method was implemented on Linux system, it is not necessarily limited to this OS and the x86-based processors. It can be adapted to environments built on multi-core processors with counting registers (linear increment) and kernels capable of binding packet processing threads to specific cores (affinity).

ACKNOWLEDGMENT

The publication was supported by the TAMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4.A/2-11-1-2012-0001 'National Excellence Program'.

REFERENCES

[1] TSC, Intel64 and IA-32 Architectures SoftwareDeveloper's Manual. Online. Available from: http://download.intel.com/products/processor/manual/325462.pdf, [retrieved: august, 2013]

[2] Linux kernel source: Linux/arch/x86/include/asm/timer.h. Online. Available from: http://www.kernel.org/, [retrieved: august, 2013]

[3] P. Orosz and T. Skopko, "Software-based Packet Capturing with High Precision Timestamping for Linux," 5th International Conference on Systems and Networks Communications, August 22-27, 2010, Nice, France, Proceeding pp. 381-386.

[4] P. Orosz, T. Skopko, and J. Imrek, "Performance Evaluation of the Nanosecond Resolution Timestamping Feature of the Enhanced Libpcap," 6th International Conference on Systems and Networks Communications, ICSNC 2011, October 23-28, 2011, Barcelona, Spain, ISBN 978-1-61208-166-3, Proceeding pp. 220-225.

[5] P. Orosz and T. Skopko, "Performance Evaluation of a High Precision Software-based Timestamping Solution for Network Monitoring," the International Journal on Advances in Software, ISSN 1942-2628, 2011 Vol 4. No. 1 & 2 pp. 181-188.

[6] P. Orosz, T. Skopko, and J. Imrek, "A NetFPGA-based Network Monitoring System with Multi-layer Timestamping: Rnetprobe," NETWORKS 2012, 15th International Telecommunications Network Strategy and Planning Symposium, October 15-18, 2012, Rome, Italy, Proceeding pp. 1-6.

[7] J. Micheel, S. Donnelly, and I. Graham, "Precision timestamping of network packets," 1st ACM SIGCOMM Workshop on Internet Measurement, November 1-2, 2001, San Francisco, California, USA, Proceeding pp. 273-277.

[8] G. Iannaccone, C. Diot, I. Graham, and N. McKeown, "Monitoring very high speed links," 1st ACM SIGCOMM Workshop on Internet Measurement, November 1-2, 2001, San Francisco, California, USA, Proceeding pp. 267-271.

[9] J. Coppens, E.P. Markatos, J. Novotny, M. Polychronakis, V. Smotlacha, and S. Ubik, "SCAMPI - A Scaleable Monitoring Platform for the Internet," 2nd International Workshop on Inter-Domain Performance and Simulation (IPS 2004), Budapest, Hungary, 22-23 March 2004

[10] A.A. Heyde, "Investigating the performance of Endace DAG monitoring hardware and Intel NICs in the context of Lawful Interception," CAIA Technical Report 080222A, august 2008.

[11] A. Pásztor and D. Veitch, "PC Based Precision Timing Without GPS," ACM SIGMETRICS 2002, Proceeding pp. 1-10.

[12] Linux NAPI device driver packet processing framework. Online. Available from: http://www.linuxfoundation.org/collaborate/workgroups/networking/napi, [retrieved: august, 2013]

 ISBN: 978-1-61208-305-6

Performance Analysis of Network Subsystem on Virtual Desktop Infrastructure System utilizing SR-IOV NIC

Soo-Cheol Oh and SeongWoon Kim
Software Research Laboratory
Electronics and Telecommunications Research Institute
Daejeon, South Korea
e-mail: {ponylife, ksw}@etri.re.kr

***Abstract*—Virtual Desktop Infrastructure (VDI) is used to run desktop operating systems and applications inside Virtual Machines (VM) that reside on servers. This paper proposes a solution for improving network performance of the VDI system. TCP/IP Offload Engine (TOE) and Single Root IO Virtualization Network Interface Card (SR-IOV NIC) for VDI protocol network and VM network were adopted respectively. According to experiments, our system saves up to 15.93% host CPU utilization.**

Keywords-Virtual Desktop Intrastructure; Virtual Network; SR-IOV NIC; PCI Passthrough; TOE

I. INTRODUCTION

Virtual Desktop Infrastructure (VDI) [1] is used to run desktop operating systems and applications inside Virtual Machines (VM) that reside on servers. The desktop operating systems inside the virtual machines are referred as the virtual desktops. Users access the virtual desktops using VDI client through network. Users can use a thin client or a zero client or a PC client. The client receives screen data of VM and displays it to the client display. Keyboard and mouse input of the client are captured and transmitted to the VM.

VDI service has many advantages. The first advantage is centralized administration that reduces maintenance cost of user desktop. The second advantage is security issue. Because all data transmission between the VM and the VDI client are controlled by an administrator, data hacking by malicious users is prohibited. The last thing is fast data backup, restoration, and provision of the virtual desktop.

However, VDI server has some disadvantages. Much load could be imposed on network subsystem and server CPU because all services are performed through the network. The screen data of the VM is a component generating much load in the VDI system. Today, popular screen size of desktop is 1920x1080p (FULLHD) with 32bit color and 60Hz refresh rate. This screen generates 3.7Gbps data. When considering blu-ray having FULLHD resolution, the compressed blu-ray data has about 61Mbps stream rate.

There are two kinds of networks in the VDI system. The first one is a VDI protocol network. Screen data from the VM to the VDI client, and input data from the VDI client and the VM are transferred through this network. The second one is a VM network that is used for intranet or internet by the VM. The VDI protocol handling large screen data of the VM generates much load on server CPU and the VDI protocol network. Also, emulation of the VM network by the server CPU generates heavy overhead on the server CPU.

The purpose of this paper is to reduce the overhead of the VDI protocol network and the VM network. For reduction of the VDI protocol network overhead, we adopt TOE [2]. Also, we reduced the VM network overhead by using SR-IOV NIC [3] and PCI passthrough technique [4].

This paper is organized as follows. Section II shows related works. Section III proposes a VDI system with improved network performance and section IV shows experimental results. Finally, section V presents our conclusions.

II. RELATED WORKS

There are several VDI systems supplied by VMware[5], Citrix[6], Microsoft[7], and KVM[8]. The VDI systems are combination of the hypervisor and the VDI protocol. The representative VDI protocols are PCoIP, ICA/HDX, and RDP/RemoteFX.

In the VDI system, Network Interface Card (NIC) for the virtual desktop is emulated by the hypervisor in software, and this emulated NIC is called as vNIC (Virtual NIC). Figure 1 shows the architecture of vNIC emulation. In this paper, we will call the NIC, physically installed on the server board, as pNIC (Physical NIC).

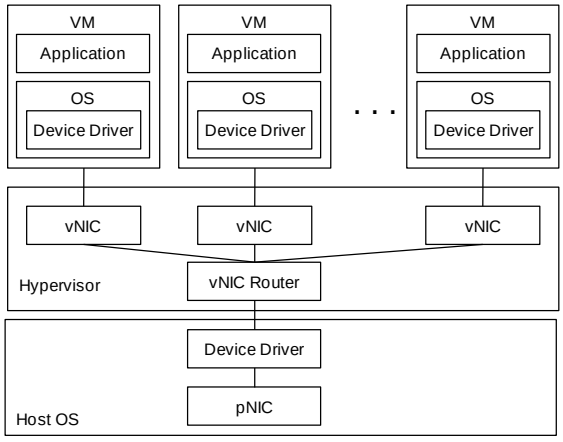


Figure 1. Software Emulation of vNIC

The hypervisor generates the vNIC based on the pNIC and provides it to the VMs. When an application on the VM wants to send data, packets including TCP/IP header and the

 ISBN: 978-1-61208-305-6

data are made by the VM. Then, these packets are delivered to the vNIC through the device driver of the VM. The hypervisor emulating the vNIC sends this packet to network using the pNIC.

When the pNIC receives data, the hypervisor decides which vNIC handles the data. Then, the hypervisor delivers this data to the vNIC. When the vNIC receives the data, it is passed to an application or an OS kernel through the device driver of the VM. The representative methods for the vNIC emulation are Network Address Translation (NAT) and bridged mode [9].

The vNIC generated by the NAT mode has private IP address and any IP connection to the outside world will look like it came from the host (same MAC and IP as the host). The network packet sent out by the VM is received by the hypervisor, which extracts the TCP/IP data, change the IP address to the IP address of the host machine, and resends it using the host OS. The hypervisor listens for replies to the packets sent, and repacks and resends them to the VM on its private network. The bridged mode means that VM will have its own MAC address and separate IP on the network and can be seen as a unique machine.

VirtIO[10], which is based on para-virtualization[11], appeared to overcome this problem. VirtIO is a virtualization standard for network and disk device drivers where the VM's device driver knows it is running in a virtual environment, and cooperates with the hypervisor. This enables the VM to get high performance network and disk operations, and gives most of the performance benefits of para-virtualization.

III. VDI System with Improved Network Performance

This paper proposes the architecture of the VDI system with improved network performance as shown in Figure 2.

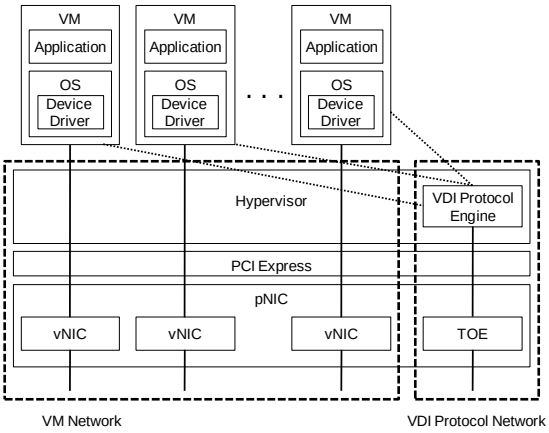


Figure 2. Architecture of the VDI System with Improved Network Performance

The VDI system is based on KVM, which is an open software. This system consists of the hypervisor, multiple VMs, and pNIC including the TOE and multiple vNICs. The hypervisor runs multiple VMs in the VDI system. The VDI protocol engine of the hypervisor sends screen data of the VM to client and receives keyboard/mouse input from the client. The pNIC installed in PCI express slot of server mother board handles network functions that are the VM network and the VDI protocol network. The VM uses the vNIC generated by the pNIC for accessing internet or intranet. The vNIC uses the SR-IOV and the PCI passthrough for performance improvement of the VM network. The VDI protocol engine uses the TOE function of the pNIC for using the VDI protocol network.

A. VDI Protocol Network

The VDI protocol network is used to deliver a virtual desktop on a server to a client. This network is managed by the hypervisor running on host OS. The data handled in the VDI protocol are summarized in Table I.

TABLE I. VDI Protocol

Data	Flow
Screen data of VM	VM --> Client
Keyboard	Client --> VM
Mouse	Client --> VM
USB	VM <--> Client

In the data, the screen data is the largest. When a VM plays a blu-ray video movie with FULLHD resolution, its average stream rate is 61Mbps. If a VDI server has 40 VMs and they play the blu-ray video movie concurrently, total network stream rate is 2.44Gbps. This high stream rate imposes high load on the host CPU for processing TCP/IP protocol. This paper adopts the TOE to solve this problem.

In the TOE, the TCP/IP protocol processing is handled in hardware NIC instead of the host CPU and this technology can reduce the load of the host CPU. Thus, we can remove the host CPU overhead for processing the VDI protocol by using the TOE.

B. VM Network

The VM network is used for accessing internet or intranet by the VM. In the previous work, the VM used the vNIC that is emulated in software by the hypervisor. This method generates heavy load on the host CPU for vNIC emulation. Also, software emulated vNIC can't fully utilize the physical performance of the pNIC because of the host CPU overhead. Although the VirtIO appeared to overcome this problem, it can't solve limitation of the software emulated NIC. This paper adopts the SR-IOV NIC and the PCI passthrough to solve this problem.

Multiple vNICs can be generated in single pNIC by utilizing the SR-IOV. Because the vNIC is emulated in hardware instead of software, it imposes no load on the host CPU.

The vNIC is attached directly to the VM using the PCI passthrough. Generally, the hypervisor manages all hardware resources and provides the virtual hardware to the VM. The PCI passthrough assigns the hardware resource to the VM directly without intervention of the host OS or the hypervisor. Thus, the VM can use the hardware resource directly without help of the hypervisor and it can reduce the overhead of the host CPU. Also, it can reduce time spent in accessing the hardware resource.

 ISBN: 978-1-61208-305-6

IV. PERFORMANCE EVALUATION

We measured performance of the network subsystem of the VDI system proposed in this paper. One server machine was connected to one client machine using 10Gigabit Ethernet Switch that is Cisco Catalyst 4900M. The specifications of the server, the client, and VM are shown in Table II.

TABLE II. CONFIGURATIONS OF SERVER, CLIENT AND VM

(a) Server Configuration

	Descriptions
CPU	- Two Intel Xeon E5-2560 2.0GHz - Total 16 cores (Each CPU has 8 cores)
Memory	128GB
Network Card	Chelsio T440-CR
Host OS	Centos 6.3 64bit

(b) Client Configuration

	Descriptions
CPU	- One Intel i7 2.67GHz - Total 4 cores
Memory	4GB
Network Card	Chelsio T440-CR
Host OS	Centos 6.3 64bit

(c) VM Configuration

	Descriptions
CPU	One virtual core
Memory	2GB
Host OS	RedHat 6.2 64bit

The Chelsio T440-CR is a 10Gbps Ethernet NIC having the SR-IOV and the TOE functions. It has four physical 10Gbps Ethernet ports. It can generate 16 vNICs per physical port. Thus, total number of the vNIC is 64.

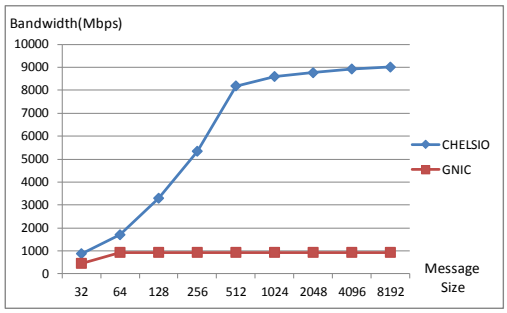
For performance comparison, we used general 1Gbps Ethernet NIC (GNIC) mounted on server main board. The experimental results are average of 10000 times.

A. The VDI Protocol Network

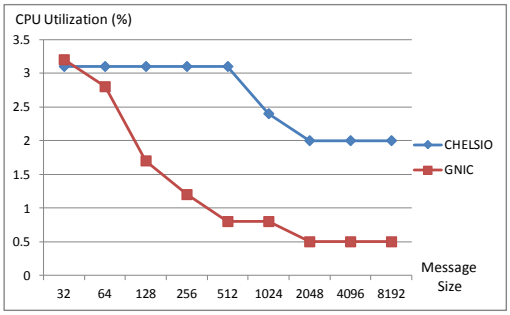
This section shows the performance comparison of the VDI protocol network. Figure 3-(a) and 3-(b) show the network bandwidth and the host CPU utilization. In this experiment, the TOE function of the CHELSIO is activated. The CHELSIO has the higher bandwidth than the GNIC. The bandwidth of the CHELSIO is 9Gbps and the GNIC is 0.93Gbps in the message size 8KB. The bandwidth of the CHELSIO increases continuously as the message size increases. However, the bandwidth of the GNIC saturates in the message size 64 bytes because the physical bandwidth of the GNIC is 1Gbps that is very lower than the CHELSIO having 10Gbps physical bandwidth.

In the host CPU utilization, the CHELSIO has the higher CPU utilization than the GNIC. The host CPU utilization of the CHELSIO with 9Gbps bandwidth is 2% in the message size 8KB. The GNIC with 0.93Gbps bandwidth consumes 0.5% CPU utilization.

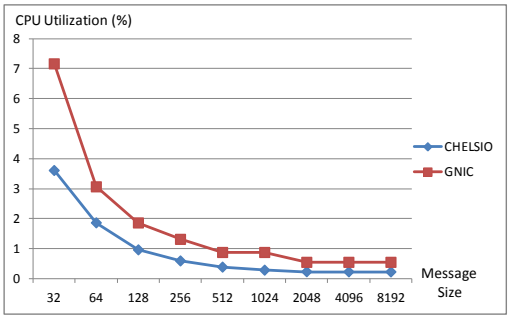
The goal of this paper is not to increase the VDI network bandwidth, but to decrease the CPU utilization of the VDI network by replacing the GNIC with the CHELSIO. In the VDI system, total data size transmitted through the VDI network is not changed although the CHELSIO sends more data than the GNIC. Thus, it is necessary to analyze the CPU utilization when sending a fixed data size. This paper used the fixed data size of 1Gbps for the analysis.



(a) Bandwidth



(b) Host CPU Utilization



(C) Host CPU Utilization for sending 1Gbps

Figure 3. Performance of the VDI protocol Network

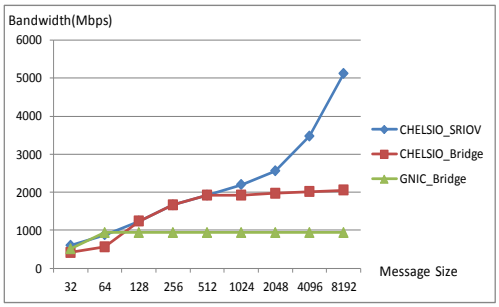
For this, we normalized the host CPU utilization and Figure 3-(c) shows the host CPU utilization spent in sending 1Gbps. To send 1Gbps, the CHELSIO consumes less CPU cycle than the GNIC. The CPU utilization of the CHELSIO is 0.22 % and the GNIC is 0.54% in the message size 8KB. Consequently, the CPU utilization of the CHELSIO is 60% less than that of the GNIC. Thus, the VDI system with the CHELSIO consumes less CPU utilization than the GNIC when sending the VDI protocol data. The saved CPU resource can be used to execute other VDI system components. The normalized CPU utilization will be used again in section IV-E.

B. The VM Network on One VM

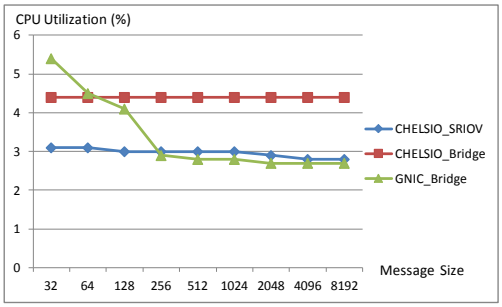
In this section, the performance of the VM network on one VM is measured. In this experiment, only one VM is

 ISBN: 978-1-61208-305-6

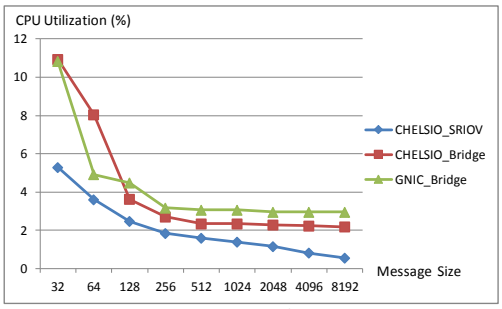
used. For performance comparison, we used three cases. The first case, i.e., CHELSIO_SRIOV, is a combination of the CHELSIO, the SR-IOV NIC, and the PCI passthrough. The second case, i.e., CHELSIO_Bridge, is combination of the CHELSIO and the bridged mode. In this case, the SRIOV and the PCI passthrough of the CHELSIO are not used. Thus, the function of the CHELSIO is the same as a 10Gbps GNIC. The last case GNIC_Bridge is combination of GNIC, the bridged mode, and VirtIO.



(a) Bandwidth



(b) Host CPU Utilization

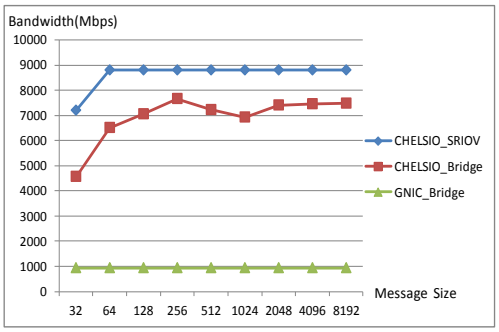


(C) Host CPU Utilization for sending 1Gbps

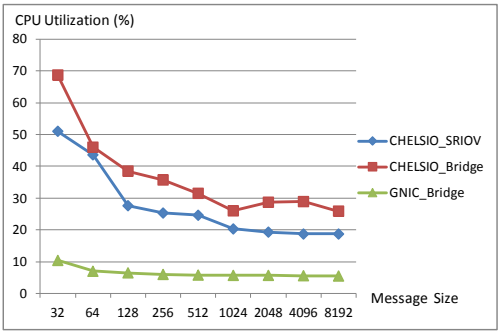
Figure 4. Performance of the VM Network on One VM

As shown in Figure 4-(a), the bandwidth values are reached, in the descending order, for: the CHELSIO_SRIOV, the CHELSIO_Bridge, and the GNIC_Bridge, respectively. The bandwidths of the CHELSIO_SRIOV and the CHELSIO_Bridge are nearly same to the message size 512 bytes. However, as the message size increases, the bandwidth of the CHELSIO_SRIOV increases to 5.12Gbps although the bandwidth of the CHELSIO_Bridge saturates at 2Gbps. The VM network of the CHELSIO_Bridge, which is the software emulated NIC, has performance limit although the physical bandwidth of the CHELSIO is 10Gbps. The GNIC_Bridge shows the 0.936Gbps that saturates for a message size of 64 bytes.

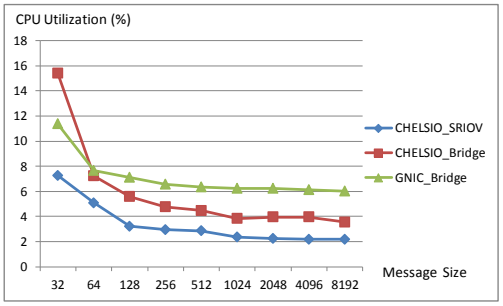
We normalized the host CPU utilization of this experiment. In all cases, the host CPU utilization decreases as the message size increases. In the message size 8KB, the host CPU utilizations of the CHELSIO_SRIOV, the CHELSIO_Bridge and the GNIC_Bridge are 0.56, 2.19 and 2.95, respectively. The CHELSIO_SRIOV has 391% and 526% less CPU utilization than the CHELSIO_Bridge and the GNIC_Bridge, respectively when sending 1Gbps.



(a) Bandwidth



(b) Host CPU Utilization



(C) Host CPU Utilization for sending 1Gbps

Figure 5. Performance of the VM Network on 16 VMs

When comparing the CHELSIO_SRIOV and the CHELSIO_Bridge, two cases use same network card. The CHELSIO_Bridge consumes more CPU cycles to emulate the bridged NIC. However, the CHELSIO_SRIOV doesn't use any CPU cycle because there is no NIC emulation. Thus, a value of 1.63% (=2.19 - 0.56) is obtained, which is the CPU utilization difference of two cases is the host CPU utilization needed to emulate the vNIC.

 ISBN: 978-1-61208-305-6

C. The VM Network on 16 VMs

The Chelsio network card generates 16 vNICs per pNIC. Thus, we measured the network performance of 16 VMs. 16 VMs are built on one server and they send data concurrently to the clients.

Figure 5-(a) shows that the CHELSIO_SRIOV has the highest bandwidth. In this experiment, the bandwidths saturate in smaller message size than section IV-B. The bandwidth saturation points of the CHELSIO_SRIOV, the CHELSIO_Bridge, and the GNIC_Bridge are obtained for 64 bytes, 256 bytes, and 32 bytes, respectively.

The normalized CPU utilizations of the CHELSIO_SRIOV, the CHELSIO_Bridge, and the GNIC_Bridge are 2.17, 3.54, and 6.01, respectively.

D. The Number VMs on one VDI Server

This section shows how many VMs are run on one VDI server. In this experiment, VMs play HD movie and the CPU utilization of each VM becomes 100%.

TABLE III. HOST CPU UTILIZATION TO 40 VMS

The number of VM	The host CPU utilization
10	43 - 46 %
20	70 - 74 %
30	91 - 92 %
40	95 - 98%

Table III shows the host CPU utilization as the number of VM increases. The host CPU utilization reaches almost 100% when the number of VM becomes 40. Thus, this VDI server can run maximum 40 VMs.

E. 100Mbps VM Network on 40 VMs

The experiments presented in sections IV-A to IV-C measured the maximum network performance of the VDI system. In this section, the network performance of the VDI system is measured in more real environment. In Korea, 100Mbps internet line to home is very popular. So, it is supposed that

- Each VM has 100Mbps VM network line.
- One VDI server supports 40 VMs.

TABLE IV. HOST CPU UTILIZATION WITH 100MBPS VM NETWORK

	CHELSIO_SRIOV	GNIC_Bridge
CPU Utilization	8.68%	24.04%

Table IV shows the host CPU utilization when each VM sends 100Mbps data concurrently (40 VMs send total 4Gbps data). The CHELSIO_SRIOV saves the 15.36% (= 24.04 - 8.68) CPU utilization compared with the GNIC_Bridge. To run one VM, 2.5% (= 100% / 40VMs) host CPU utilization is needed. Thus, saved 15.35% host CPU utilization can be used to run 6 more VMs (= 15.35% / 2.5%).

F. The Movie Play

Let's consider that 40 VMs connect Youtube website and play FullHD movie (1920x1080p). We used Gangnam style music video of PSY. The service scenario of this experiment is as follow:

1) Compressed FullHD movie of the Youtube site is streamed to VM using the VM network.

2) The VM decodes the compressed movie and shows decompressed movie on screen of the VM.

3) The VDI protocol engine shown in Figure 2 delivers this screen data to the client using the VDI protocol network.

This section analyzes the scenario by utilizing the experimental results of section IV-A to IV-D, and shows how much host CPU utilization is saved when using CHELSIO_SRIOV.

The specification of the PSY music video is as follow:

- S_{VMNET} : 1.6Gbit
  - o Size of compressed Youtube music video delivered through the VM network (unit : Gbit)
- S_{VDINET} : 11.1Gbit
  - o Size of the VM screen data delivered through the VDI protocol network (unit : Gbit)
- T_{play} : 4 minutes 13 seconds = 253 seconds
  - o Play time of the music video (unit : second)

Equation (1) shows the host CPU utilization (U_{total}) consumed in the network subsystem of the VDI server.

$$U_{total} = \sum_{i=1}^{N} U_{VDINET\ i} + \sum_{i=1}^{N} U_{VMNET\ i} \quad (1)$$

- U_{total} : host CPU utilization consumed in the network subsystem of the VDI server
- N : the number of VM
- $U_{VDINETi}$: host CPU utilization consumed in the VDI protocol network of VM_i
- U_{VMNETi} : host CPU utilization consumed in the VM network of VM_i

Because we assume that all VMs use same video, assumptions of equation (2) and (3) can be applied to (1), and U_{total} is expressed in (4) and (5).

$$U_{VDINET} = U_{VDINET\ 1} = U_{VDINET\ 2} = \cdots = U_{VDINET\ N} \quad (2)$$

$$U_{VMNET} = U_{VMNET\ 1} = U_{VMNET\ 2} = \cdots = U_{VMNET\ N} \quad (3)$$

$$U_{total} = N \times U_{VDINET} + N \times U_{VMNET} \quad (4)$$

$$U_{total} = N \times U_{VDINET\ \_1G} \times \frac{S_{VDINET}}{T_{play}} + N \times U_{VMNET\ \_1G} \times \frac{S_{VMNET}}{T_{buffer}} \quad (5)$$

- U_{VDINET} : average host CPU utilization consumed in the VDI network per VM
- U_{VMNET} : average host CPU utilization consumed in the VM network per VM

 ISBN: 978-1-61208-305-6

- U_{VDINET_1G} : host CPU utilization consumed for transferring 1Gbit data through the VDI protocol network
- U_{VMNET_1G} : host CPU utilization consumed for transferring 1Gbit data through the VM network
- T_{buffer} : time spent in buffering all music video through the VM network per VM

According to sections IV-A to IV-C, the parameters of the CHELSIO_SRIOV and the GNIC are as follow:

- CHELSIO_SRIOV
  - o U_{VDINET_1G} (0.22727), U_{VMNET_1G} (2.176)
- GNIC
  - o U_{VDINET_1G} (0.54701), U_{VMNET_1G} (6.017)

When the VM receives the music video from the Youtube site, buffering is utilized. The size of the PSY music video is 1.6Gbit and the VM has 100Mbps VM network. The earliest time in that the buffering finishes is 16 seconds. If the internet has bad quality, the time could be extended to the movie play time that is 253 seconds.

Table V shows the average host CPU utilization when buffering through the VM network finishes within a specified time. This table is derived from (5).

TABLE V. HOST CPU UTILIZATION ON 40 VMS

Time	VM Network			VDI Network			Diff Sum
	CHELSIO_SRIOV	GNIC	Diff	CHELSIO_SRIOV	GNIC	Diff	
16	8.7	24.07	15.37	0.4	0.96	0.56	15.93
20	6.96	19.25	12.29	0.4	0.96	0.56	12.85
40	3.48	9.63	6.15	0.4	0.96	0.56	6.71
60	2.32	6.42	4.1	0.4	0.96	0.56	4.66
80	1.74	4.81	3.07	0.4	0.96	0.56	3.63
100	1.39	3.85	2.46	0.4	0.96	0.56	3.02
120	1.16	3.21	2.05	0.4	0.96	0.56	2.61
140	0.99	2.75	1.76	0.4	0.96	0.56	2.32
160	0.87	2.41	1.54	0.4	0.96	0.56	2.1
180	0.77	2.14	1.37	0.4	0.96	0.56	1.93
200	0.7	1.93	1.23	0.4	0.96	0.56	1.79
220	0.63	1.75	1.12	0.4	0.96	0.56	1.68
240	0.58	1.6	1.02	0.4	0.96	0.56	1.58
253	0.55	1.52	0.97	0.4	0.96	0.56	1.53

Let us consider the case where the buffering finishes in 16 seconds. Average 8.7% host CPU utilization is consumed to use the VM network in the CHELSIO_SRIOV. However, the GNIC uses 24.07% host CPU utilization. Thus, the CHELSIO_SRIOV saves the 15.35% host CPU utilization for the VM network. In the VDI network, the CHELSIO_SRIOV saves the 0.56% host CPU utilization.

Consequently, in time zone 0 to 16 seconds, the CHELSIO_SRIOV save average 15.93% host CPU utilization than the GNIC.

After the buffering is finished, the VM network is used no more and the benefit of the CHELSIO_SRIOV for the VM network is 0. In time zone 17 to 253 seconds, the benefit of the CHELSIO_SRIOV is 0.56% host CPU utilization that is from the VDI network.

As the buffering finishes in longer time, the benefit of the CHELSIO_SRIOV decreases. If the buffering finish time becomes 253 seconds, the CHELSIO_SRIOV can save 1.53% host CPU utilization than the GNIC.

According to the experiment in Korean internet environment, buffering finishes in 60 seconds to 120 seconds. Thus, the CHELSIO_SRIOV saves 4.17% to 7.38% host CPU utilization.

V. CONCLUSIONS

This paper proposed a solution for improving network performance of the VDI system. The TOE and the SR-IOV for VDI protocol network and VM network were adopted respectively. This removed the host CPU overhead used to emulate the vNIC in software by the hypervisor. Also, we removed the host CPU overhead consumed to process the VDI protocol network by using TOE. According to experiments, our system saved up to 15.93% host CPU utilization.

VI. ACKNOWLEDGEMENT

This work was supported by the IT R&D program of MSIP/KEIT. [10035242, Development of Cloud DaaS (Desktop as a Service) System and Terminal Technology]

REFERENCES

[1] D.-A. Dasilva, L. Liu, N. Bessis, and Y. Zhan, "Enabling Green IT through Building a Virtual Desktop Infrastructure", 2012 Eighth International Conference on Knowledge and Grids (SKG), Beijing, Oct. 2012, pp. 32-38.

[2] S.-C. Oh and S. W. Kim, "An Efficient Linux Kernel Module supporting TCP/IP Offload Engine on Grid", Fifth International Conference on Grid and Cooperative Computing, Hunan, Oct. 2006, pp. 228-235.

[3] C.H. N. Reddy, "Hardware Based I/O Virtualization Technologies for Hypervisors, Configurations and Advantages - A Study", 2012 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bangalore, Oct. 2012, pp. 1-5.

[4] M. T. Jones, "Linux virtualization and PCI passthrough", IBM Developer Works Technical Library, Oct. 2009, http://www.ibm.com/developerworks/library/l-pci-passthrough, [retrieved: Oct. 2013].

[5] J. Langone and A. Leibovici, "Chapter5. The PCoIP Protocol", VMware View 5 Desktop Virtualization Solutions, Jun. 2012, pp. 77-87.

[6] G. R. James, "Chapter 5. Desktop Delivery Controller", Citrix XenDesktop Implementation, 2010, pp. 113-127.

[7] T. Cerling, J. Buller, C. Enstall, and R. Ruiz, "Chapter 15. Deploying Microsoft VDI", Mastering Microsoft Virtualization, Nov. 2011, pp. 443-476.

[8] I. Habib, "Virtualization with KVM", Linux Journal, Volume 2008, Issue 166, Feb. 2008. Article No. 8.

[9] Virtualbox, "Chapter 6. Virtual Networking", Oracle VM VirtualBox User Manual, http://www.virtualbox.org/manual/ch06.html, [retrieved: Oct. 2013].

[10] R. Russell, "virtio: towards a de-facto standard for virtual I/O devices", ACM SIGOPS Operating Systems Review - Research and developments in the Linux kernel, Volume 42, Issue 5, Jul. 2008, pp. 95-103.

[11] VMware, "Understanding Full Virtualization, Paravirtualization, and Hardware Assist", VMware Technical Resource Center Technical Papers, Nov. 2007.

 ISBN: 978-1-61208-305-6

Improving Reliability of Inter-connected Networks through Connecting Structure

Yuka Takeshita, Shin'ichi Arakawa, and Masayuki Murata
Graduate School of Information Science and Technology
Osaka University, Japan
{y-takeshita, arakawa, murata}@ist.osaka-u.ac.jp

***Abstract*—The Internet plays an important role in our life as social infrastructure, and the importance of reliability is widely recognized in the Internet. There are many studies on network design with high reliability but most of them intend for constructing a single network that a network operator governs. However, the Internet consists of many of small networks, which are mutually connected. Therefore, it is important to enhance reliability of inter-connected network consisted from two or more networks rather than focusing only on the reliability inside the single network. In this paper, we show how we should connect two networks for achieving high reliability of inter-connected network. We evaluate the reliability with various kinds of connecting structures. Evaluation results show that high reliability is achieved by a multiscale structure where links for inter-connection are prepared for connecting nodes belonging to different hierarchical level in the network.**

***Keywords*—Power-law Networks; BA Model; Reliability; Connecting Structure; Multiscale Structure.**

I. INTRODUCTION

The number of users connected to the Internet is increasing through mobile terminals and various services such as social networking service are deployed. The Internet plays an important role in our life as social infrastructure, and therefore reliability is one of the important characteristics for the Internet.

Internet Service Providers (ISPs) construct their own networks to accommodate the traffic of customers with a minimum of equipment costs while keeping the reliability against failures of equipment [1]. A key functionality to keep the reliability is the restoration, i.e., re-route packets when failures occur. Network operator of ISP envisions kind of failures and then designs physical topology and capacity of links so that the network works under the envisioned failures. However, when more significant failures than initially envisioned, the network becomes out of control, that is, it may work or may not work. The network operator faces on the difficulty in deciding the scale of failures of undertakings.

In previous studies, a single node failure and/or single link failure were supposed as the failure of equipment [2], [3]. The fundamental approach of these studies is to enumerate all of failure patterns and then prepares physical links or determines the capacity of links to accommodate traffic demand for all of failure patterns. However, it is easily imagined that such the approach encounters the difficulty in designing networks when the size of simultaneous failures envisioned increases. The reliability against multiple node/link failures is investigated in [4], [5]. They focus on the statistical characteristics of topology and investigate the relation between the characteristics and the reliability under the multiple failures. Results show that the power-law network where the probability of existence of nodes having k links is proportional to $k^{-\gamma}$ (γ is constant) loses its connectivity easily when nodes with high degree are failed, but the power-law network is reliable against random node failures.

Above studies intend for enhancing reliability of an ISP network that the network operator governs. However, the reliability of the Internet is achieved not only by the enhancing reliability of ISP networks but also by enhancing reliability of inter-connected network, where two or more ISP networks are mutually connected, since the Internet consists of many of ISPs which are mutually connected. In this paper, we investigate the reliability of inter-connected network that consists from two networks and their connecting links. Hereafter, we will call the inter-connected network as global network, and call its consisting networks as local networks. Our concern is how we should connect a limited number of inter-connected links between local networks to make the global network to be reliable against multiple failures. Note that we evaluate connecting structure between local networks rather than the topological structure of local network itself, since the reliability of local network has been investigated in the above studies.

Recently, the reliability of electronic network that consists from power-grid network and its control network is discussed [6]–[8]. Since the control network requires the power from the power-grid network, the authors investigate that how to inter-connect two networks such that reliability against cascade failures is maximized. The cascade failure is successive failures caused by a cascade of power-outage which is triggered by an initial failure point. They pointed out that the global network is reliable against cascade failures when two local networks are connected with links through "similar" nodes. That is, inter-connected links should be prepared between nodes with similar degree or similar clustering coefficient. Unlike the electronic network where nodes of control network must be connected with the power-grid network, communication network does not require full connectivity between two networks. Rather, it is important for communication networks to reduce the number of inter-connected links to keep the reliability to some extent. Note again that our concern of this paper is how we should connect a limited number of inter-connected links between local networks, which is particular to communication networks.

This paper is organized as follows. We introduce related work of this paper in Section II. Section III shows the topology model that we use for the evaluations. In Section IV, we evaluate reliability with various classes of connecting structure against node failures. Finally, Section V concludes this paper and mentions the future work.

 ISBN: 978-1-61208-305-6

II. Related Work

Dodds et. al. showed a network construction algorithm that constructs five classes of networks and compared their robustness [4]. The algorithm starts from a hierarchical tree topology with branching ratio b and L levels of branching. Then, the algorithm adds m links chosen stochastically with a probability. The probability that there exists a link between two nodes, say i and j, is denoted as $P(i,j)$ and is determined by the depth D_{ij} of their nearest common ancestor a_{ij}. The probability is also determined by node's own depths d_i and d_j (Fig. 1). Formally the probability $P(i,j)$ is defined as,

$$P(i,j) \quad \propto \quad e^{-D_{ij}/\lambda} e^{-x_{ij}/\zeta}, \tag{1}$$

where λ and ζ are tunable parameters. x_{ij} represents the distance between two nodes i and j and is set to $(d_i^2 + d_j^2 - 2)^{1/2}$, which represents relative distance in the hierarchy [4]. By changing the values of λ and ζ, this algorithm generates topologies with various topological structures. The authors categorized generated topologies into the following five classes.

- Random (R) by setting $(\lambda, \zeta) \rightarrow (\infty, \infty)$: links are added randomly.
- Random interdivisional (RID) by setting $(\lambda, \zeta) \rightarrow (0, \infty)$: more links are added for smaller value of D_{ij}, but do not take care of x_{ij}. That is, the link between nodes that have large distance.
- Local Team (LT) by setting $(\lambda, \zeta) \rightarrow (\infty, 0)$: links tend to be added between nodes that have short distance, regardless of their layer in hierarchy.
- Core-periphery (CP) by setting $(\lambda, \zeta) \rightarrow (0, 0)$: links tend to be added between nodes located at higher-level in hierarchy, and between nodes that have short distance. The resulting topology exhibits densely connected "core" and sparsely connected "edge" network.
- Multiscale (MS) with intermediate values of λ ($0 < \lambda < 1$) and ζ ($0 < \zeta < 1$). The resulting topology has connectivity dominated by the range from a small x_{ij} to a large x_{ij}. The resulting topology has a property that the link density decreases as the hierarchical level decreases.

The authors evaluated two kinds of robustness. One is congestion robustness and the other is connectivity robustness. Congestion robustness is measured by the maximum congestion that imposes a load of packet processing at node. Connectivity robustness represents the size of the largest connected component remaining after failures. Their evaluation reveals that the multiscale structure improves both the congestion robustness and connectivity robustness.

III. Connecting structure for inter-connected network

In this section, we present a model of connecting structure between two local networks inspired by the network construction algorithm explained in the previous section. There are two local networks: local network A and local network B. The global network (inter-connected network) is formed by connecting links between A and B. Depending on a strategy where to connect, various connecting structure can be arranged.

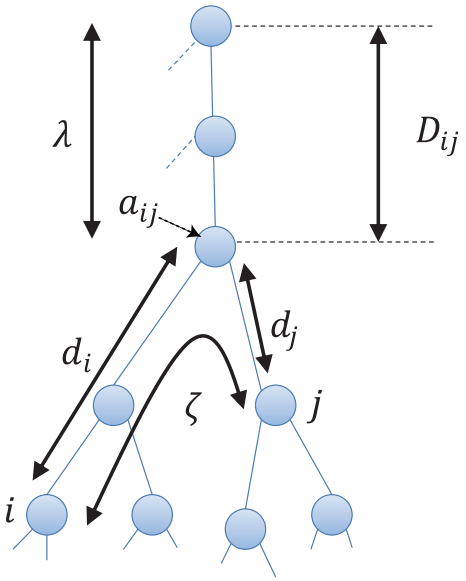


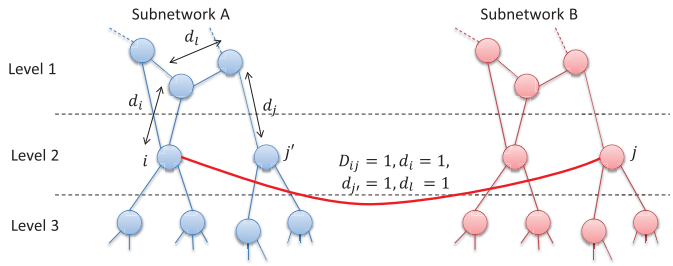
Fig. 1. Illustrative example of the network construction algorithm ($D_{ij} = 2$, $d_i = 2$, and $d_j = 1$)

For developing the model, we assume that two local networks are identical. Note that such the assumption does not reflect the actual network. However, we use the assumption in this paper since our main concern is to reveal fundamental reliability of inter-connected network and investigate differences of the reliability on various connecting structures. Actually, the reliability may be different dependent on things of each local network. We will consider networks having different topology as a future work. We also assume that the local network has a hierarchy structure and has a level of hierarchy.

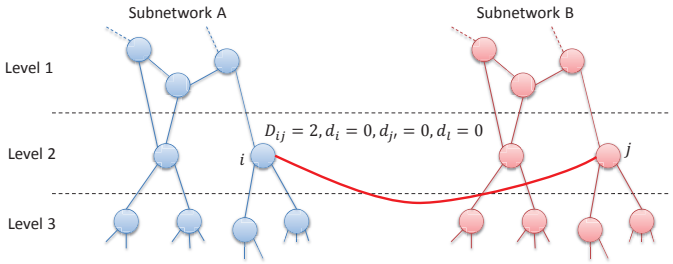
Let us consider that the probability $P(i,j)$ which represent the probability of link existence between node i from local network A and node j from local network B. Then, we calculate connection probability $P(i,j)$ of all nodes pairs (i,j), which is defined as,

$$P(i,j) \propto e^{-D_{ij}/\lambda} e^{-x_{ij}/\zeta}. \tag{1}$$

Note that this equation is the same to the equation in [4]. However, we change the definition of each notation to apply our problem that connects two local networks. First, we redefine the distance x_{ij} by using three values d_i, d_j, and d_l. Hereafter, we use a node j' of local network A instead of a node j of local network B. Node j' of local network A corresponds to the node j of local network B. Note again that we assume that local networks A and B are identical to reveal the reliability of global network. In our model, d_i is defined as the number of upstream hops in the shortest path from source node i to a common ancestor $a_{ij'}$. Similarly, we define d_j as the number of downstream hops from a common ancestor $a'_{ij'}$ to node j. d_l represents horizontal distance in the hierarchical local network and is set to the shortest hop length between i and j excluding d_i and d_j. In this paper, we introduce a concept of horizontal distance to consider a non-tree-based topology as the local network. Dodds et al. [4] consider the tree-based topologies for network construction and the non-tree-based topology is not treated. Illustrative example of d_i,



(a) $i \neq j'$



(b) $i = j'$

Fig. 2. Definition of x_{ij} and D_{ij} used for connecting two local networks

TABLE I. VALUES OF (λ, ζ)

notation of connecting structure	(λ, ζ)
Random (R)	(∞, ∞)
Local Team (LT)	$(\infty, 0.05)$
Random Interdivisional (RID)	$(0.05, \infty)$
Core-periphery (CP)	(0.05, 0.05)
Multiscale (MS)	(0.5, 0.5)

d_j, and d_l, is shown in Fig. 2. Then, the distance x_{ij} is re-defined as $(d_i^2 + d_j^2 + d_l^2)^{1/2}$.

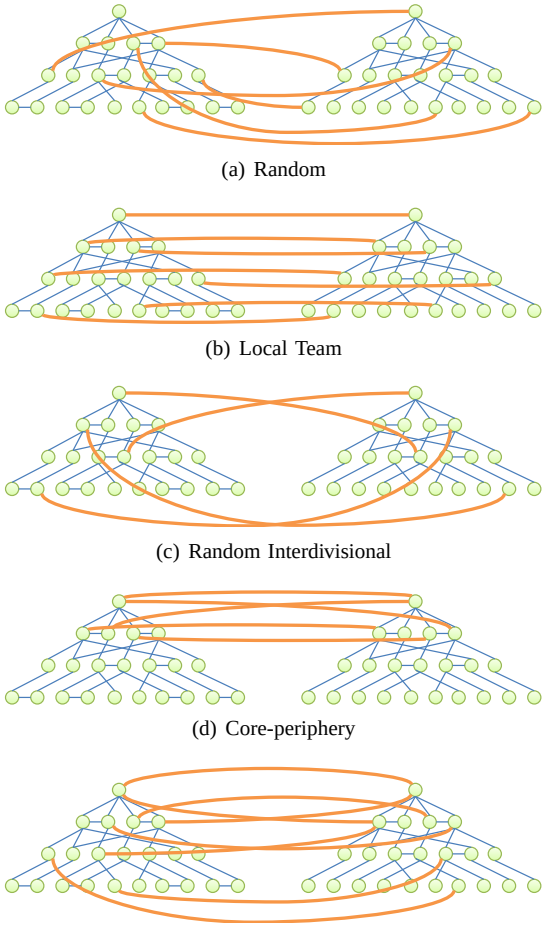
After calculating connection probability, we connect two nodes belonged to different local networks. We select connected nodes pair i and j according to $P(i, j)$. We then connect between node i in local network A and node j in local network B. We repeat adding links between two local networks until the number of inter-connected links reaches m.

By changing the parameters λ and ζ, we generate some classes of inter-connected topology. We use the same definition of classes in the way of [4] (shown in Table I and Fig. 3). However, Multiscale structure is defined as the middle parameters of other four structures, so we cannot set the unique value for Multiscale structure. Therefore, we evaluate some parameters other than $(\lambda, \zeta) = (0.5, 0.5)$. We set the number of inter-connected links to 50, 100, and 200.

IV. RELIABILITY EVALUATION OF INTER-CONNECTED NETWORK

A. Local Network

We prepare a local network based on BA model [9]. BA model is a well-known generation model for topology whose degree distribution follows a power-law. The BA model incrementally adds a new node, and the new node connects with existing nodes by a preferential manner, i.e., new nodes tends to connect higher degree node. The detailed of algorithm to generate the BA topology is as follows:



(a) Random

(b) Local Team

(c) Random Interdivisional

(d) Core-periphery

(e) Multiscale

Fig. 3. Five classes of connecting structure obtained by changing parameters, λ or ζ.

1) Prepare a complete graph with m_0 nodes
2) Repeat following processes until the number of nodes equal to n
   a) set a new node
   b) select m nodes with the probability $k_i / \sum_j k_j$ (k_i is the degree of node i) and connect between selected nodes and a new node.

In this paper, we consider four patterns of local network by changing values of (n, m) to (500, 2), (500, 3), (1000, 2), (1000, 3). m_0 is set to 3 for all patterns. Hierarchical level of BA topology is defined by the hop count from the node with largest degree in the local network.

B. Performance Metrics

We evaluate the average hop length and the connectivity when multiple failures occur. Hereafter, N denotes the number of nodes and B denotes the largest connected component after the failures occur.

- Average hop length H
  H denotes the average hop length for all pairs of

nodes, which is defined as

$$H = \frac{\sum_{i \in B}^{N} \sum_{j \neq i, j \in B}^{N} d_{ij}}{|B|(|B|-1)}, \quad (2)$$

where d_{ij} is the shortest hop length from node i to node j calculated by Dijkstra's shortest path algorithm.

- Connectivity C
  C denotes the ratio of the number of nodes in B to a set of all survived nodes, which is defined as

$$C = \frac{|B|}{N - |r|}, \quad (3)$$

where r is a set of failed nodes. $|B|$ and $|r|$ means the number of elements in each set.

C. Reliability against node Failures

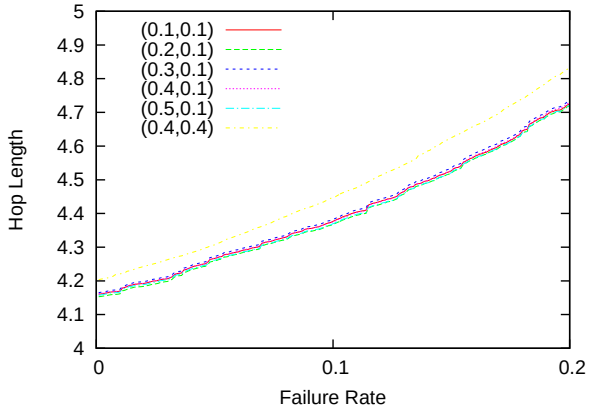
In this section, we consider the scenario that a node failure occurs at random one by one.

1) Parameter settings for Multiscale Structure: As we discussed in Section II, Multiscale structure is intermediate of other four structures (Random, Local Team, Random Interdivisional, Core-periphery). Since the parameters λ and ζ takes various values, we first investigate the best values of the parameters for multiple node failures. In [4], setting λ to 0.5 and $zeta$ to 0.5 exhibits best parameter setting for improving robustness for constructing a local network. A question of this paper is whether setting λ to 0.5 and ζ to 0.5 is best or not.

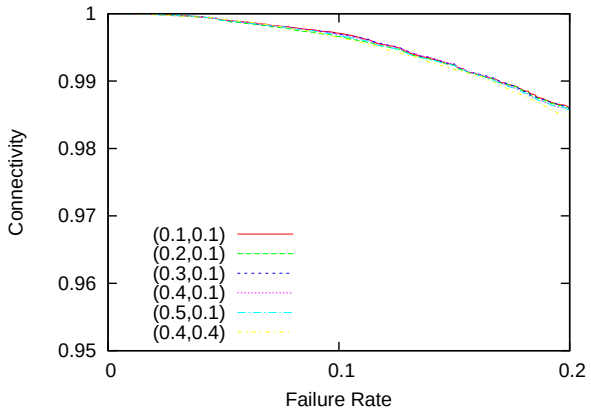
In [4], congestion robustness is improved when the Multiscale structure close to the Core-periphery structure. We therefore investigate the parameter set which is close to Core-periphery structure. More specifically, we evaluate reliability by changing λ and ζ from 0.1 to 0.5 by 0.1, respectively. We calculate average of C and H for 100 patterns of local networks having 500 nodes with average degree 2. The number of inter-connected links is set to 200.

For obtaining best parameter settings, we change ζ from 0 to 0.5 while λ is fixed. When λ is set to 0.1, 0.1 for $zeta$ exhibits best reliability in terms of connectivity and average hop count. We next set λ to 0.2, but again 0.1 for ζ exhibits best reliability. Finally, we obtained that setting ζ to 0.1 is best for λ 0.1, 0.2, 0.3, 0.5, except when λ is set to 0.4; setting ζ to 0.4 exhibits best reliability. Figure 4 shows connectivity C and average hop-count H for different value of λ. For each value of λ, ζ is chosen such that the reliability is maximized. Comparing results with various λ, we observe that the average hop length is minimized when ($of \lambda$, ζ) is (0.1, 0.1), but this is so close to Core-periphery structure that we do not select the parameters as Multiscale structure. Though we cannot see notable differences on the average hop length and the connectivity, we select is (0.3, 0.1) as the parameter (λ, ζ) since the connectivity C is slightly higher than other parameters.

The reason for showing high reliability is that MS (0.3, 0.1) has more inter-connected links that connect nodes at 2nd layer than MS (0.5, 0.5). To clarify this, we show the number of nodes in each layer in Table II, and the number of inter-connected links dependent on layers in the hierarchy in Table III (MS (0.5, 0.5)) and Table IV (MS (0.3, 0.1)). When (λ, ζ)



(a) average hop length



(b) connectivity

Fig. 4. reliability of topology formed by (λ, ζ) = (*, 0.1), (0.4, 0.4)

is set to (0.3, 0.1), the number of inter-connected links related to 2nd layer is 72, which is larger than the case of MS (0.5, 0.5). In the case of BA topology, nodes at 2nd layer tend to connect with node at 1st layer, i.e., largest degree node. The average hop length is therefore decreased for MS (0.3, 0.1).

From these observations, we investigate MS (0.3,0.1), which we set λ to 0.3 and ζ to 0.1, as well as MS (0.5, 0.5) for Multiscale structure.

TABLE II. THE NUMBER OF NODES IN EACH LAYER: 500 NODES, AVERAGE DEGREE 2 FOR LOCAL NETWORK

Hierarchical level	number of nodes
1	1
2	61
3	169
4	228
5	41

TABLE III. LAYERS WITH INTER-CONNECTED LINKS OF NETWORKS BY ($\lambda = 0.5$, $\zeta = 0.5$)

Hierarchical level	1	2	3	4	5
1	0	0	0	2	0
2	0	6	8	12	3
3	0	9	23	28	8
4	1	10	27	45	10
5	0	1	2	3	2

 ISBN: 978-1-61208-305-6

TABLE IV. LAYERS WITH INTER-CONNECTED LINKS OF NETWORKS BY ($\lambda = 0.3, \zeta = 0.1$)

Hierarchical level	1	2	3	4	5
1	0	0	0	0	0
2	1	10	21	12	1
3	0	17	29	20	6
4	0	9	28	33	2
5	0	1	5	4	1

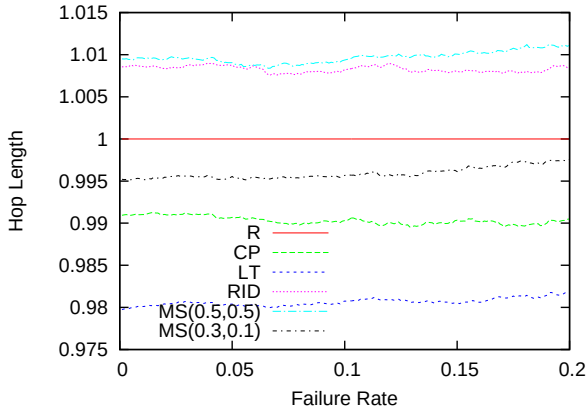


Fig. 5. Average hop length for multiple failures: 500 nodes, average degree 2 for local networks; 50 inter-connected links.

2) Evaluation on Connecting Structure: We evaluate reliability of networks with MS (0.3, 0.1) in addition to five classes of connecting structures shown in Table I. We show the average hop length in Fig. 5 with 500 nodes and average degree 2 for local networks. In this figure, X-axis shows the ratio of node failures and Y-axis shows average hop length normalized by the result of Random structure. We observe that the structures with dense links in upper layers, such as Core-periphery structure or Local Team structure, could make the average hop length to be low. However, when we change the number of nodes or links in local networks, average hop length H of Core-periphery structure and Local Team structure get worse, and sometimes close to that of Random structure. MS (0.3, 0.1) can keep the average hop length low regardless of the number of nodes or links used for local networks.

Next, we show the connectivity C in Fig. 6. In this figure, X-axis shows the ratio of node failures and Y-axis shows connectivity C normalized by the result of Random structure. We can see that MS (0.3, 0.1) or MS (0.5, 0.5) show higher connectivity than that of other structures.

We also show worst case of connectivity C and average hop length H in Figs. 7 and 8. The definition of X-axis and Y-axis is the same to the definition of Fig. 6. As shown in Fig. 7, we cannot observe any remarkable differences among results of each connecting structure. This is because we use BA topology as local networks. BA topology has degree distribution obeying a power-law and the topology already has a robustness against random node failures. However, Fig. 8 shows that MS (0.3,0.1) can take higher connectivity C than other structures against failure rate. These results show that MS (0.3,0.1) showed low average hop length and high connectivity when multiple node failures occur.

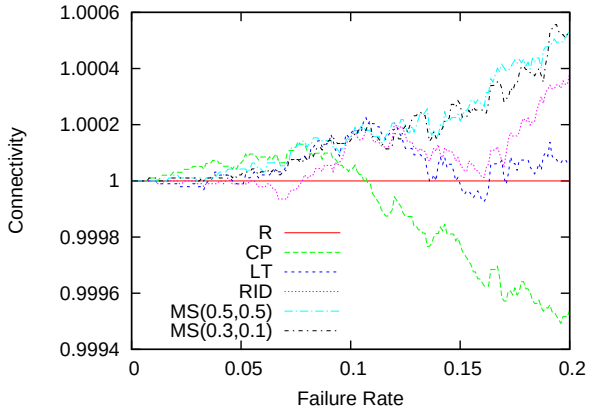


Fig. 6. Connectivity for multiple failures: 500 nodes, average degree 2 for local networks; 50 inter-connected links.

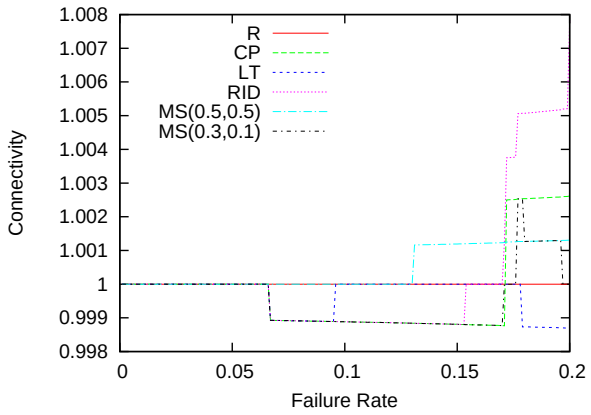


Fig. 7. Worst case of connectivity C for multiple node failures: 500 nodes, average degree 2 for local network; 50 inter-connected links

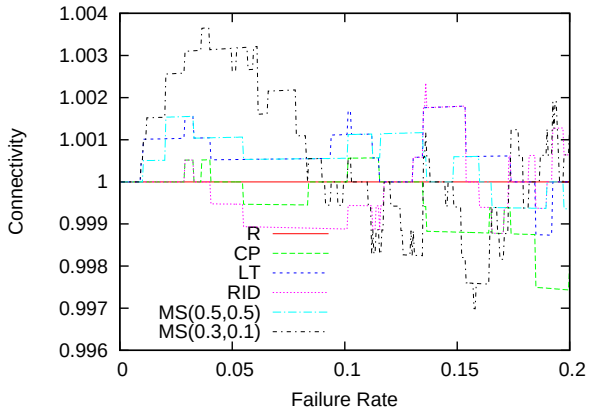


Fig. 8. Worst case of connectivity C for multiple node failures: 1000 nodes, average degree 2 for local network; 200 inter-connected links

 ISBN: 978-1-61208-305-6

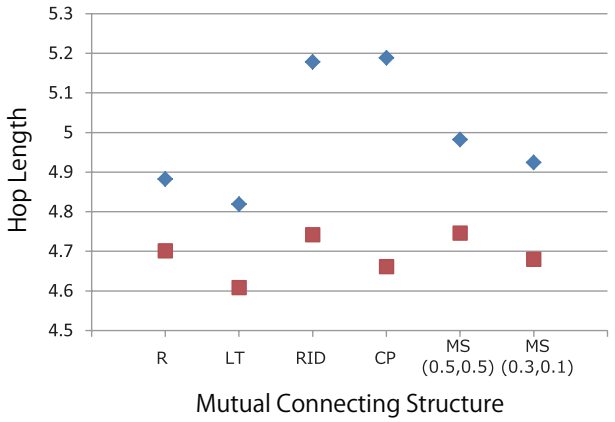


Fig. 9. The average hop length for disaster failure: 500 nodes, average degree 2; 50 inter-connected links). ■ shows the average and ◆ shows the worst values

D. Reliability for Disaster Failures

In the previous section, we evaluated reliability against multiple node failures where a single node failure successively and randomly occurs one by one. This section evaluates reliability of inter-connected network against disaster failure. As opposed to random node failures examined at previous section, we consider multiple node failures where a selected node and its neighbor nodes fail simultaneously. For evaluating the reliability against disaster failure, we consider failures of largest degree node and its neighbor nodes. This is the worst case scenario for the disaster failure since the scale of disaster is largest. Of course, it is possible to occur multiple disasters at the same time, but the possibility is extremely low, so we do not evaluate multiple disaster scenario.

We examined various local networks by changing the parameter for generating BA topology. Figure 9 is the results when we use 500-node with average degree 2 as the local network. In this figure, X-axis represents classes of connecting structure, and Y-axis represents the average hop length H. Worst and average of H over 100 patterns of local network is presented for each class of connecting structure. The results show that MS (0.3, 0.1) and Local Team structure can keep the average hop length low for both worst and averaged results. It is also revealed that Core-periphery structure and Random Interdivisional structure takes high average hop length at the worst case. We had the same observation for other local networks, so omitted results in this paper. We also omit evaluations of connectivity C since we cannot see remarkable differences among connecting structures.

Based on these results, we conclude that MS (0.3, 0.1) structure shows high reliability for multiple node failures and a disaster failure. That is, high reliability of inter-connected is achieved by connecting nodes belonged to different hierarchical level in local network and by connecting nodes around the core of local network densely.

V. Conclusion and Future Work

In this paper, we revealed how we should connect two local networks for achieving high reliability of inter-connected network. For this purpose, we extend the algorithm in [4] with re-definition of distance x_{ij} between nodes i and j. We then examined various classes of connecting structures between two local networks, and evaluate the connectivity and average hop length after multiple node failures. The results showed that high reliability is achieved by MS (0.3, 0.1), which is the Multiscale structure with λ 0.3 and ζ 0.1. The other structures sometimes take high reliability, but MS (0.3, 0.1) always takes high reliability.

In the future work, we will investigate the reliability of inter-connected network between two ISP topologies other than BA topologies, and extend the definition of the probability $P(i,j)$ to be applied to connect two local network whose topologies are different from each other.

Acknowledgment

This work was supported in part by Grant-in-Aid for Scientific Research (A) 24240010 of the Japan Society for the Promotion of Science (JSPS) in Japan.

References

[1] G. Iannaccone, C.-N. Chuah, S. Bhattacharyya, and C. Diot, “Feasibility of IP restoration in a tier 1 backbone,” IEEE Network, vol. 18, no. 2, pp. 13–19, Aug. 2004.

[2] L. Shen, X. Yang, and B. Ramamurthy, “Shared risk link group (SRLG)-diverse path provisioning under hybrid service level agreements in wavelength-routed optical mesh networks,” IEEE/ACM Transactions on Networking, vol. 13, no. 4, pp. 918–931, Aug. 2005.

[3] A. Hansen, A. Kvalbein, T. Čičić, and S. Gjessing, “Resilient routing layers for network disaster planning,” Lecture notes in computer science, vol. 3421, pp. 1097–1105, Apr. 2005.

[4] P. S. Dodds, D. J.Watts, and C. F.Sabel, “Information exchange and the robustness of organizational networks,” in Proceedings of the National Academy of Sciences (PNAS), vol. 100, Oct. 2003, pp. 12516–12521.

[5] X. Wang and G. Chen, “Complex networks: small-world, scale-free and beyond,” IEEE Circuits and Systems Magazine, vol. 3, pp. 6–20, Jan. 2003.

[6] R. Parshani, C. Rozenblat, D. Ietri, C. Ducruet, and S. Havlin, “Inter-similarity between coupled networks,” Europhysics Letters, vol. 92, pp. 68002(1)–68002(5), Jan. 2011.

[7] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, “Networks formed from interdependent networks,” Nature Physics, vol. 8, pp. 40–48, Dec. 2011.

[8] C. D. Brummitt, R. M. D’Souza, and E. A. Leicht, “Suppressing cascades of load in interdependent networks,” in Proceedings of the National Academy of Sciences (PNAS), vol. 109, Mar. 2012, pp. 681–689.

[9] A. Barabási and R. Albert, “Emergence of scaling in random networks,” Science, vol. 286, pp. 509–512, Oct. 1999.

 ISBN: 978-1-61208-305-6

Toward Reliability Guarantee VC Services in an Advance Reservation based Network Resource Provisioning System

H. Lim
Dept. of High Energy Physics
California Institute of Technology
Pasadena, CA, USA
hklim@caltech.edu

Y. Lee
Dept. of Computer Engineering
Mokpo National University
Mokpo, Jeon-Nam, South Korea
yhlee@mokpo.ac.kr

***Abstract*— The most representative research network in Korea, KREONET, has developed DynamicKL, an advance reservation based network service agent for user driven virtual circuit services. DynamicKL provides reservation, provisioning, release, termination, and inquiry web services for network resources by using an open standard network service interface (NSI), as well as web services for network resources by using a GUI interface. In addition, it has the RESTful web service Interface for Configuration and Event management (RICE) to support a protection management function for virtual circuits and reservations. In this paper, a protection management per virtual circuit (VC) for provisioned VCs and reservations is addressed in the DynamicKL framework, as a contribution to the VC protection management issue, which results in more manageable and reliable VC services compared to other advance reservation frameworks. An administrator can detect successful or unsuccessful VC protections in the event of a primary link failure and successful or unsuccessful VC retrievals after a primary link repair, by using RICE.**

Keywords- Advance reservation, Network Service Agent (NSA), Network Service Interface (NSI), Dynamic provisioned network resource, DynamicKL, Virtual circuit protection, Primary link, Secondary link, Link failure.

I. INTRODUCTION

Recently, most dynamic provisioning in advanced research networks have developed and deployed advance reservation based network resource provisioning systems for big data transfers to support various application areas, for example, DRAGON, OSCARS, DRAC, AutoBHAN, EnLIGHTened, PHOSPHOROUS, and G-Lambda, [1][2][3][4][5][6][7]. They have their own framework for only network resources or for both grid and network services. Some of them have a standard interface for multi-domain services and the others have their own interface.

The Network Service Interface (NSI) developed by the Open Grid Forum (OGF) is a standard interface for network resource reservation and control in intra or inter-domain [8][9][10]. An NSI based resource reservation and provisioning system can improve productivity for data intensive research projects, for example, reserving and allocating available network resources (i.e., virtual circuits) automatically for large-scale data applications between multi-domains, such as Large Hadron Collider (LHC) in the field of High Energy Physics (HEP). G-Lambda A/K (the latest version of G-Lambda), OSCARS, AutoBHAN, and OpenDRAC (the latest version of DRAC) have been implementing the standard NSI interface in their frameworks [11].

Dynamically provisioned network resources, such as VCs, are recognized as extremely useful capabilities for many types of network services. However, to date the majority of approaches to such services do not address protection management per VC to protect data traffics in VCs in the event of node or link failures, which provide manageability and reliability guarantees VC services in advance reservation frameworks.

We present Dynamic circuit based advance reservation system of KRLight (DynamicKL) based on web services, which consists of the Network Service Agent (NSA) and a web portal server. In particular, DyamicKL provides the RESTful web service Interface for Configuration and Event management (RICE) web service interface for the protection management per VC for virtual circuits (VCs) and reservations in the event of a link failure, as well as the NSI and Graphic User Interface (GUI) web services for reservation, provisioning, release, termination, and inquiry. Protection management in DynamicKL is provided per VC for provisioned VCs and reservations in case of a link failure, a feature that contributes to managing failure and protection status information per VC in a primary and backup VC reservation DB. This capability constitutes a dominant, important difference from other advance reservation systems. With this capability, an administrator can detect when backup VCs are successfully or unsuccessfully working as active paths to protect primary VCs and primary VCs are successfully or unsuccessfully retrieved as active paths after a primary link repair. Because a primary and backup reservation DB separately manage failure information per each primary/ backup VC, it is possible to establish VCs and to terminate reservations in backup links for provisioning and termination requests of reservations with a primary link failure, by delivering NSI provisioning messages with backup interface information.

In Section II, other advance reservation frameworks are introduced as related works. The DynamicKL framework with protection management is addressed in Section III. In Section IV, protection management function per VC is addressed to provide more manageable and reliable VC

 ISBN: 978-1-61208-305-6

services more detail, which differentiates this approach from other advance reservation frameworks. To verify its capabilities, a protection management demonstration for virtual circuits with a link failure is presented in Section V. Finally Section VI gives conclusions and implications.

II. Related Works For Advance Reservation Framework

In this section, we introduce architecture and development of other advance reservation frameworks selected as representative related research. They are mainly focused on VC service technology issues other than management issues such as protection management per VC, for reliability and manageability guarantee VC services to protect data traffics to disjointed VCs in the event of a link or node failure.

A. DRAGON

Dynamic Resource Allocation via GMPLS Optical Networks (DRAGON) [1] is a project that allows dedicated network resources dynamically to link computational clusters, storage arrays, and other instruments into distributed topologies. GMPLS [12] is used to create virtual circuits for both optical and Ethernet domains and DRAGON creates Layer 1 and Layer 2 virtual circuits. The key components of DRAGON software consist of Virtual Label Switched Router (VLSR), Network Aware Resource Broker (NARB), Application Specific Topology Builder (ASTB), and Resource Computation Engine (RCE). To create virtual circuits that cover various domains, the NARB acts as the entity that represents a local autonomous system or a domain [1]. The main role of VLSR is to control Ethernet switches via the GMPLS control plane. RCE and ASTB are used to compute network resources needed for virtual circuit provisioning.

B. OSCARS

On-Demand Secure Service and Advance Reservation System (OSCARS) is a user driven network software developed to support dynamic virtual circuits (VCs), for large scale data applications such as Large Hadron Collider (LHC) research using the Energy Science Network (ESnet) in the US [2][13]. The main objective of OSCARS is to allow application programmers and end-users to set up advance reservations for VCs [2]. MPLS-TE and RSVP-TE protocols are used to make advance reservations and to allocate dedicated bandwidth on demand. The Label Switched Paths (LSPs) are created for Layer 2 and Layer 3 circuits using OSCARS software [2][13]. The key components of OSCARS consist of the Reservation Manager (RM), Path Setup Subsystem (PSS), Bandwidth Scheduler Subsystem (BSS), Authentication, Authorization and Accounting Subsystem (AAA) [2]. OSCARS is currently implementing open standard interface (NSI) in its framework.

C. AutoBHAN

AutoBAHN is a GEANT-provided provisioning tool that integrates with an NREN's own systems to facilitate the multi-domain dynamic circuit provisioning service [3]. AutoBAHN eliminates the long established problem of manually provisioning multi-domain circuits, reducing this time from weeks to a matter of minutes, even seconds. AutoBAHN easily negotiates the different networking technologies deployed by the different domains. Interoperability with other BoD systems is achieved through the Inter Domain Controller Protocol (IDCP), and the Network Services Interface (NSI) Protocol, fully enabling global connections.

The AutoBAHN system is based on the Inter-Domain Manager (IDM), a module responsible for the inter-domain operation of circuit reservation on behalf of a domain [3].

D. DRAC

The Dynamic Resource Allocation Controller (DRAC) [4] is a network service to support network resources automatically and dynamically, to meet application requirements in SURFnet (the national education and research network in the Netherlands). DRAC acts as an agent of the various applications, brokering and configuring on an end-to-end basis all the necessary pieces of the network, regardless of the type of network – circuit or packet, wireless or wired network [4]. DRAC allows very large number of flows of packets or low-latency applications dynamically through Layer 1 circuit instead of Layer 3. DRAC simply create and release optical circuits as application requirements [4].

DRAC has been extended to OpenDRAC. Open DRAC is an open source project that is developing a middleware that allows network control by users and applications [13][14]. It is currently compatible with open standard interface (NSI).

E. G-Lambda

G-Lambda [7] is a project to provide users with virtual dedicated circuits for both grid and network resources in Japan. The GNS-WSI [15] in G-LAMBDA defines a set of messages to be sent to Network Resource Manager (NRM) from Grid Resource Scheduler (GRS). These include messages to reserve, activate, release, and inquire VCs.

The G-Lambda framework consists of a GRS, which behaves as a Grid Scheduler, and Resource Managers (NRM for network resources and CRM for computing resources), which manage each local network or computing resource. GRS and RMs work together to provide users with co-allocation and resource reservations. The GRS provides a web service interface to user clients using the web service resource framework (WSRF) [7][15]. The NRM is responsible for path virtualization between endpoints, local scheduling, and activation/de-activation of VCs [7][13][15]. G-Lambda is currently extended to G-Lambda-A/K with open standard interface (NSI).

 ISBN: 978-1-61208-305-6

III. DYNAMICKL FRAMEWORK WITH PROTECTION MANAGEMENT

A. DynamicKL System Block

DynamicKL consists of a web portal server and a NSA, as shown in Fig. 1. The web portal server provides a web-based user interface for users to make advance reservations. It has a primary and backup VC reservation DB, a network topology DB, and a user account DB, and provides an AAA module for the basic user authentication and authorization process. Primary and backup reservation DBs separately manage failure and protection status information per each primary and backup VC. The NSA system consists of an NSI Handler (NSIH) with a Provider Agent (PA) and an Requester Agent (RA) to support network resource service in intra or inter domain, a Path Computation and Resource Admission (PCRA) and a G-UNI Message Handler (GUMH) to manage network resource in intra domain, and a Configuration and Event Management Handler (CEMH), as shown in Fig. 1. The web portal server interfaces with the NSIH through the Network Service Interface (NSI) and the CEMH through the RICE, respectively.

The NSIH executes advance reservations based connection management in intra or inter-domain with the NSI interface. Also, the PA in the NSIH interfaces with the PCRA through the GNSI interface for connection management for intra domain network resources. The PA delivers requested reservation information to the PCRA through the GNSI interface and the PCRA performs path computation and an admission control for local network resource reservation. The PCRA reflects node or link failure information received from the CEMH in network topology information and has a primary and backup VC reservation table managed with ResvID. By using them, the PCRA controls admission for new VC reservation request. The PCRA interfaces with the GUMH for the creation and release of virtual circuits on a network path requested by a user. The GUMH exchanges control messages for creation, release and inquiry of primary and backup virtual circuits with network devices through the GUNI interface. The GUMH receives network failure/repair events by SNMP trap messages from network devices. To detect a router (node) failure event, periodic polling messages from the GUMH are received at network devices. VC protection/retrieval events can be detected by using a *Query_VC* GUNI message from the GUMH. The CEMH provides a management plane with network event and VC protection/retrieval event information received from the GUMH through RICE API messages. The CEMH internally interfaces with the PCRA, to initialize and apply network topology information received from the web portal server, and to request renewal of network topology and backup/primary VC reservation table information. Also, the CEMH internally interfaces with the GUMH, to detect network event information from network devices and to request VC protection/retrieval information inquiry.

The NSI is a standard interface for network resource reservation and control between inter domains defined by Open Grid Forum (OGF), in partnership with the Global Lambda Integrated Facility (GLIF) organization. The Grid Network Service Interface (GNSI) is an interface between

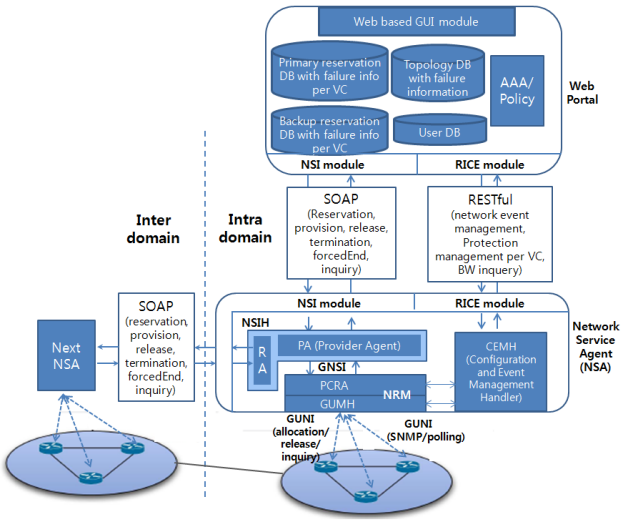


Fig. 1. DynamicKL framework with a protection management per VC.

NSIH and PCRA used for the reservation and connection management of the local domain [16]. The Grid User Network Interface (GUNI) is an interface for VC creation, release and inquiry. The RICE is an interface for network event management, especially for protection management per VC.

B. Interfaces in DynamicKL

1) NSI interface

NSI messages [8][9][10] for network resource control are shown in Table I. A request message from the RA is delivered to the PA. The web portal server plays a role of the RA for user VC services. The PA sends a confirmation or failure response message to the RA as identification of success or failure for a request message from the RA. A NSI message is delivered through Simple Object Access Protocol (SOAP). All of the NSI messages have a correlation ID as an identifier for a request and response message and a request message has a next NSA address for inter domain VC service. All of the NSI messages are defined in Table I.

TABLE I. NSI V1.1 MESSAGES FOR NETWORK RESOURCE SERVICE

Message	Description
reserveRequest	A message from RA that requests a network resource to PA for a connection between two nodes
provisonRequest	A message from RA that allocates a reserved network resource to PA
releaseRequest	A message from RA that releases an allocated network resource and maintains reservation for the network resource to PA
terminateRequest	A message from RA that terminates an allocated or a reserved network resource to PA
queryRequest	A message from RA that inquiries connection status to PA
forcedEndRequest	A message from PA that notifies RA that PA administratively terminated a reservation

 ISBN: 978-1-61208-305-6

A *reserveRequest* message is used to request a VC reservation. A reservation request message has following factors: globalReservationId, connectionId (CID), service parameter and path information. A starting time, an end time and a bandwidth are included in service parameter. Direction (bi-direction as a default), addresses of source/destination nodes are included in path information. A *reserveConf* message is used to response for a reservation request. [8][9][10]. For provisioning of a reserved VC identified as a CID, a *provisionRequest* message is used. For release of a provisioned VC, a *releaseRequest* message is used and a *releaseConf* message is used for response. Also, a *terminateRequest* message is used for termination of a reserved or a provisioned VC and a *terminateConf* message is used for a response. Finally, a *queryReqeust* message is used for inquiry of a reserved or provisioned VC and a *queryConf* message is used for response. A *forcedEndRequest* message is used to notify RA that PA administratively terminated a reservation [8][9][10]. If any service for reservation, provisioning, release, termination, and inquiry has failed, PA sends a failure message to RA.

2) GNSI interface

GNSI is an interface for a network resource reservation service defined by the GLIF organization [16]. GNSI has been previously implemented in a network resource provisioning system for local domain VC service by us [16]. We make use of it as an internal interface in DynamicKL for intra domain VC services. The GNSI messages implemented for resource reservation service is as follows. *GreateResourceResv* is used for resource reservation and *ProvisionResourceResv* is used to allocate a reserved network resource. *ReleaseResourceResv* and *ReleaseResourceProv* are used to release a reserved resource and a provisioned resource, respectively. *GetResourceProperty* is used to return attribute information corresponding to a reserved resource. The *GetAllReservedResources* message is used to inquire about an available BW between a requested reservation starting time and a requested reservation end time in a designated network path [16]. To provide interoperability between the NSI and GNSI interfaces, a CID2ResvID mapping table is used in NSIH.

3) GUNI interface

The GUNI is an interface for VC creation and release for a reserved resource. *Activate_VC* and *Deactivate_VC* messages are used to create and release primary and backup virtual circuits on a requested network path, respectively. *Query_VC* is used to inquire about secondary (backup) VCs working on active or non-active status, in the event of a primary link failure, and primary VCs working on active or non-active status, in case of a primary link repair. Each message includes information to create and release and inquire VCs by telnet access to each network device. To receive SNMP trap messages from network devices in the event of network failures, SNMP trap based GUNI interface is also applied to GUMH.

4) RICE interface

We designed and implemented the RESTful web service Interface for Configuration and Event management (RICE) for the application of network topology information to the NSA, BW inquiry for a specific path with a reservation duration, and network event management, especially for protection management per VC for provisioned VCs and reservations with a link failure. The RICE API messages for protection management per VC will be described in detail in section III.

C. Virtual Circuit Provisioning by DynamicKL

DynamicKL supports advance reservations of VCs at layer 2 (VLANs) through a NSI interface, and layer 3 (MPLS LSPs) through a NSI and GUI interface. In this capacity, DynamicKL is used as an intra domain controller for network resources within KREONET. DynamicKL also functions as a inter-domain controller which has the capability to communicate with other intra domain NSAs through NSI interface, as shown Fig. 1. DynamicKL is used to allow application programmers and end-users to set up reservations for VCs in advance, with NSI and GUI interfaces.

Once a reservation is made, a VC provisioning step can be instantiated either by the NSI provisioning request from a user. Network device specific GUNI messages for each bender are used to initiate RSVP signaling, and MPLS LSP provisioning and release on the network devices. LSPs are established based on Forward Equivalent Class (FEC) for VC provisioning between two storage hosts. For both layer 2 and layer 3 VCs, where reservations are bidirectional, the configuration GUNI messages are delivered to both edge routers at the start and end of the intra domain VC. 2 backup VCs (i.e., 1 bidirectional VC) in secondary links are internally provisioned for protection, in addition to 2 primary VCs (i.e., 1 bidirectional VC) provisioned in primary links by a user request.

D. VC Reservation Request by DynamicKL

A user can select source/destination nodes on the topology map and provide source/destination host addresses, which is internally mapped to Service Termination Point (STP) addresses. If a user provides a starting time and an end time of reservation and inquires about a residual bandwidth, a maximum available bandwidth from a source node to a destination node is shown to a user. If a user provides a bandwidth smaller than that, a reservation request is ended. So, users do not have to experience reservation failures when searching for a needed specific BW. 2 backup VCs (i.e., 1 bi-directional VC) in secondary links are internally reserved for protection by DynamicKL, in addition to 2 primary VCs (i.e., 1 bi-directional VC) in primary links by a user request.

 ISBN: 978-1-61208-305-6

Since users can notice primary link or node failure in a topology map, they can reserve VCs on the rest of nodes and primary links, except failure nodes or links.

E. Monitoring

An administrator can monitor network events and all VCs and reservations at all sites in a dynamic VC network, while a user can monitor his/her VCs or reservations only. Thus, an administrator has an authorization to terminate reservation abuses in all sites, in case of unexpected events, such as router or interface failures or suspected reservation abuses. An administrator can request a network operator in a dynamic VC network to recover physical network failures and VC failures due to unsuccessful protection or unexpected events.

IV. PROTECTION MANAGEMENT PER VC

Since NSI standardization does not yet address network management issues, such as protection management, we have implemented the RICE, especially for protection management for VCs and reservations with a link failure.

A backup virtual circuit for protection is internally reserved and provisioned by DynamicKL, together with a primary reservation and virtual circuit for a user request. An active path is switched from a primary VC to a backup VC in a dynamic VC service network, in the event of a primary link failure (i.e., a 1:1 path based protection [17] implemented in network devices is used). A protection management function per VC is addressed to provide manageable and reliable VC services, by using RICE.

A. RICE API messages for network event management

1) InterfaceDown

When an interface of a network device has a fault, an SNMP trap message from a network device is delivered to the GUMH in NSA [18]. An *InterfaceDown* message is used to notify a web portal server a fault interface of a network device.

2) InterfaceUp

When a failure interface is repaired, an SNMP trap message from a network device with a fault interface is delivered to the GUMH. An *InterfaceUP* message is used to notify a web portal server a retrieved interface of a network device.

3) NodeDown

To monitor a network device with a fault, a periodic polling message from the GUMH is delivered to network devices. A *NodeDown* message is used to notify a web portal server a fault of a network device.

4) NodeUp

When a network device with a fault is retrieved, it can be monitored by using both an SNMP trap and polling messages [18]. A *NodeUp* message is used to notify a web portal server a retrieved network device.

5) Primary2SecondarySuccess/Primary2SecondaryFail and Secondary2PrimarySuccess/Secondary2PrimaryFail

Primary2SecondarySuccess and *Primary2SecondaryFail* API messages are used to notify that backup VCs pre-assigned in secondary links currently operate as working paths to protect VCs in a failure primary link and at least a backup VC does not operate as a working path, respectively. On the other hand, *Secondary2PrimarySuccess* and *Secondary2PrimaryFail* messages are used to provide notification that all of VCs in a repaired primary link (interfaces) are retrieved as active paths, and to indicate that at least a VC is not currently retrieved as an active path, respectively. The above events can be detected by sending Query_VC GUNI messages to network devices, after receiving an SNMP trap message with primary link failure or repair information from a network device.

B. A protection management service scenario

RICE and GUNI message flows for a protection management service scenario are shown in Fig. 2. Internal message flows between the PCRA, the CEMH and GUMH are also shown. It is assumed that a failure primary link has provisioned VCs beforehand. A SNMP trap message from a network device with a failure link is received at the GUMH. The GUMH notifies the CEMH failure link information and the CEMH requests the PCRA to renew network topology and primary reservation table information. The CEMH sends an *InterfaceDownRequest* message to the web portal server for a notification of a failure link and the web server renews network topology and primary reservation DBs. An *InterfaceDownConf* message is received at the CEMH. Because provisioned VCs in a failure primary link have also failures, the CEMH requests the GUMH to inquire about status information of backup VCs. The GUMH detects status information of backup VCs by using a *Query_VC* GUNI message. If all backup VCs pre-assigned are working on active paths, the GUMH notifies the CEMH a protection success. The CEMH requests the PCRA to renew a backup

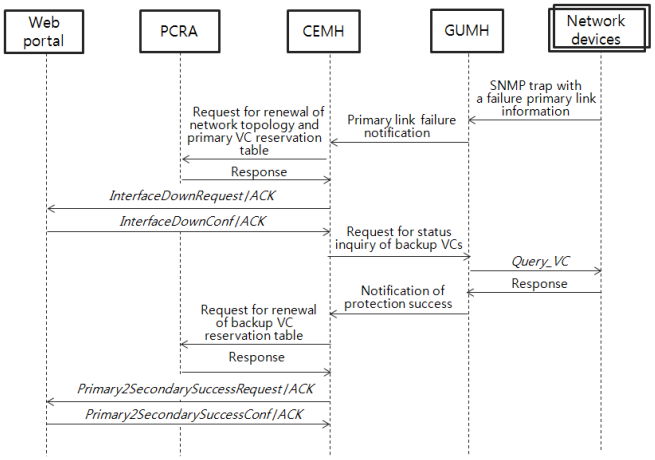


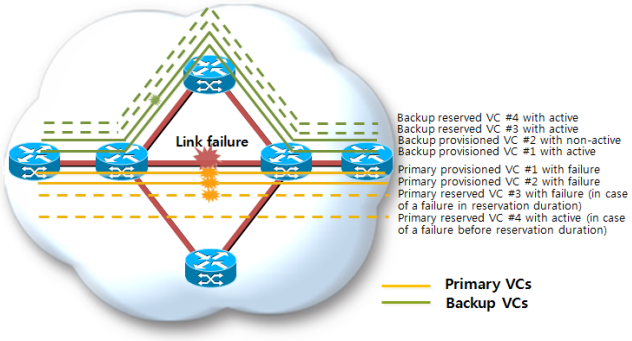
Fig. 2. RICE and GUNI message flows for one protection management service scenario.

VC reservation table. The CEMH finally sends a *Primary2SecondarySuccessRequest* message to the web portal server for a notification of a protection success. A backup VC reservation DB changes status of secondary VCs to active status. Then the web portal server sends a *Primary2SecondarySuccessConf* message to the CEMH.

An administrator can identify each network event by receiving a distinguished RICE message from NSA.

C. Advantages of Protection Management per VC

By using the RICE API messages, an administrator can observe that backup VCs are currently working as active paths or not working in the event of a primary link failure, and primary VCs are automatically retrieved or not retrieved after a primary link repair. With these messages, failure and protection status information is managed per VC, as well as per network link. In other words, NSA notifies successful or unsuccessful VC activations as working paths in the backup link and successful or unsuccessful VC retrievals as working paths after a primary link repair, and failure and protection status information per VC (i.e., per CID) in a primary and backup reservation DB is separately managed, as shown in Fig. 3. If protection status information is not provided per VC, an administrator will assume that all backup VCs are successfully operating as working paths in the event of a failure primary link with VCs and that all primary VCs operates successfully as working paths in the event of a primary link repair. We also note that a provisioned VC may not operate as an active path, due to unexpected events, such as a network device's configuration intervention by an individual and an OS problem in a network device. In addition, a provisioned VC may be released by an administrator's mistake.



<Protection management per VC in primary/backup reservation DBs>

VC	Primary VC status	Backup VC status
VC #1	Provisioned with failure	Provisioned with non-active
VC #2	Provisioned with failure	Provisioned with active
VC #3	Reserved with failure	Reserved with active
VC #4	Reserved with active	Reserved with active

Fig. 3 Protection management per VC in a primary and backup reservation DB in the event of a primary link failure with VCs and reservations.

Because primary and backup reservation DBs separately manage failure and protection status information per each primary and backup VC, as shown in Fig 3, DynamicKL can still release and terminate user virtual circuits in backup links, in the event of a primary link failure with VCs. Also, it is possible to establish VCs and to terminate reservations in backup links for reservations with a primary link failure, by delivering NSI provisioning messages with backup interfaces information to network devices. Users wanting to reserve VCs can monitor a failure link or a failure node and create reservations for the rest of links and nodes.

D. Comparison of Advance Reservation Frameworks

Table II provides a comparison of advance reservation frameworks. G-Lambda A/K, OSCARS, AutoBHAN, and OpenDRAC have been implementing the standard NSI interface in their frameworks. Few of the works in literatures have been introduced [3][11][19]. As an additional function compared to other advance reservation frameworks, DynamicKL provides a protection management function per VC, for protection from primary VCs to disjointed VCs and its specific management, in the event of a failure link.

Protection management in an advance reservation framework for VC services should be provided per VC, in protecting provisioned VCs and reservations with a link failure. That is, an advance reservation framework should be able to detect successful or unsuccessful VC activation information in backup links and successful or unsuccessful VC retrieval information in primary links. With this information, it is possible to renew and manage a primary and backup reservation DB with failure and protection status information per VC (CID), which leads to a protection management per VC. These capabilities in the DynamicKL result in more manageable and reliable VC service.

V. Demonstration of Protection Management per VC

A. Service Network Architecture

To address the growing need for guaranteed bandwidth by large-scale collaborations, such as the LHC in the field of HEP, the KREONET has designed and implemented the dynamic virtual circuit network, which is physically distinct from the IP core network (KREONET). NSI VC services (i.e., reservation, provision, release, termination, and inquiry services) can be made on the dynamic VC service network, which consists of some part of 5 sites in the KREONET, as shown in Fig. 4. The KREONET core network is architected primarily to transport IP packets as a best effort service, while the KREONET dynamic virtual circuit network is engineered to support only dynamic virtual circuits (VCs). The NSA system implemented as web service operates as a server to process request messages and operates as a client for the creation of confirmation messages. The web portal

 ISBN: 978-1-61208-305-6

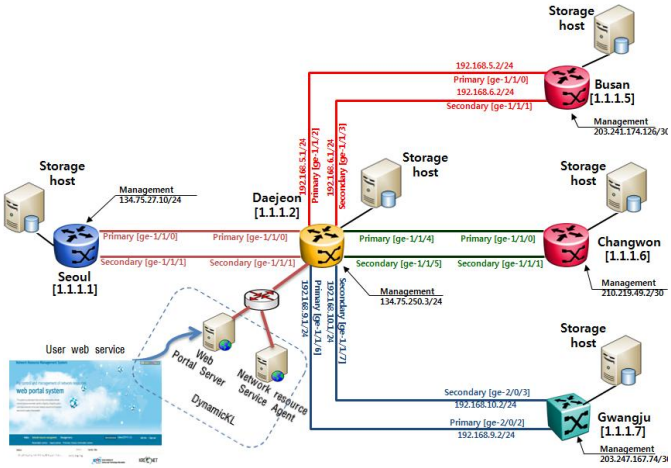


Fig. 4. Dynamic VC service network by the DynamicKL with a protection management per VC.

server also operates as a client for request messages and a server for confirmation messages.

Even though a secondary (backup) and primary VC in the dynamic VC network do not have disjointed paths in the event of a node failure, due to the star topology of KREONET core network, they have disjointed paths each other in the event of a link failure. A demonstration for protection management per VC is focused on the event of a link failure.

B. Demonstration of Protection Management per VC

In this subsection, a demonstration of protection management per VC for VCs with a link failure is demonstrated in the dynamic VC service network. Bidirectional VC (2 unidirectional VCs) with 100 *Mbps* BW on a designated network path from a host at Gwangju to a host at Pusan is provisioned by a user request. A primary link at Daejeon site has a failure event. A protection management GUI for an administrator is shown in Fig. 5. An administrator can notice that a primary link at Daejeon site has failure, by receiving an *InterfaceDown* message. Also an administrator

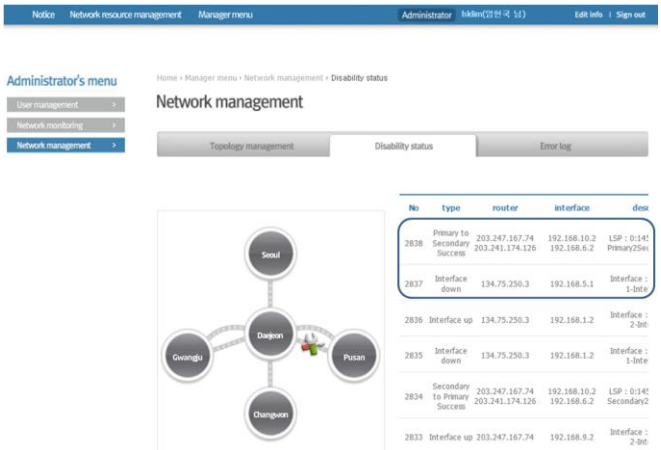


Fig. 5. Protection management GUI for an administrator.

can recognize that backup VCs in secondary links are successfully working as active paths to protect primary VCs with a link failure, by receiving a *Primary2SecondarySuccess* API message. A primary reservation DB creates failure information for primary VCs and a backup reservation DB changes status information for backup VCs from non-active (standby) status to active status. By making use of these API messages, protection management per VC is possible in a primary and backup reservation DB. Figure 6 verifies that two secondary VCs pre-assigned are working as active paths, by exchanging signaling messages between network devices after a primary link failure.

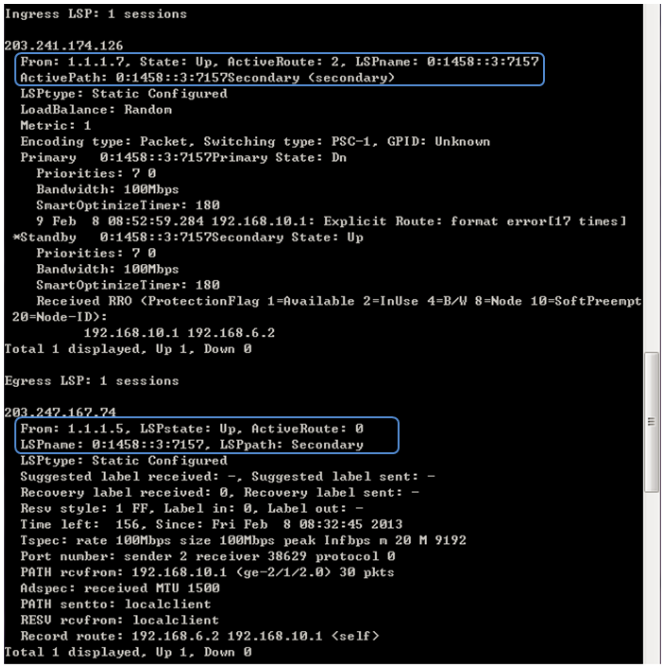


Fig. 6. Secondary VCs working on active paths in the event of a primary link failure.

When a failure primary link at Daejeon site is repaired by a network operator, a management GUI for an administrator is shown in Fig. 7. An administrator can recognize that a failure primary link was repaired by receiving an *InterfaceUp* message. Also, an administrator can notice that VCs in primary links are successfully retrieved as active (working) paths, by receiving a *Secondary2PrimarySuccess* API message. A primary reservation DB changes status information of provisioned VCs with failures to active status and a backup reservation DB renews backup VCs with active status to them with non-active (standby) status. These RICE API messages demonstrate that a retrieval management per VC is possible in DynamicKL.

VI. CONCLUSIONS

To date the majority of approaches in advance reservation frameworks do not address a number of required management issues, such as fault and protection managements for VC services, to provide manageability and

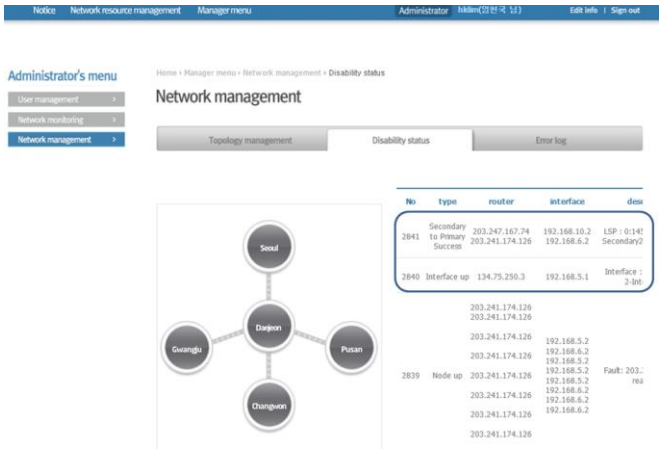


Fig. 7. VC retrieval management GUI for an administrator.

reliability guarantees. Here, we demonstrated that the DynamicKL provides the protection management per VC by using RICE interface, for virtual circuits and reservations with a primary link failure, which leads to more manageable and reliable VC services. Also, DynamicKL provides reservation, provision, release, termination, and inquiry services for virtual circuits by using a standard network service interface (NSI). With the protection management capability, an administrator can notice successful or unsuccessful VC protections in the event of a primary link failure and successful or unsuccessful primary VC retrievals as active paths after a primary link repair. Because a primary and a backup reservation DB separately manage failure and protection status information per each primary and backup VC, DynamicKL is able to release and terminate user virtual circuits in backup links, in the event of a primary link failure with VCs. In conclusion, DynamicKL could contribute to improve manageability and reliability of VC services.

REFERENCES

[1] T. Lehman, J. Sobieski, and B. Jabbari, "DRAGON: A Framework for Service Provisioning in Heterogeneous Grid Network," IEEE Communi., Magazine, Vol. 44, Issue 3, Mar. 2006, pp. 84-90.

[2] C. P. Guok, D. W. Robertson, E. Chaniotakis, M. R. Thompson, W. Johnston, and B. Tierney, "A User Driven Dynamic Circuit Network Implementation", IEEE Globe Communi., Oct. 2009, pp. 1-6.

[3] J. Lukasik, O. Neofytou, A. Sevasti, S. Thomas, and S. Tyley, "Installation and Deployment Guide: AutoBAHN System Book", Published by DANTE, June 2008.

[4] F. Travostino, R. Keates, T. Lavian, I. Monga, and B. Schofield, "Project DRAC: Creating an Application-aware Network", Nortel Journal, vol. 22 No. 8, Oct. 2006, pp. 23-26.

[5] G. Zervas, et al., "Phosphorus grid-enabled GMPLS control plane (GMPLS): architectures, services, and interfaces," IEEE Communication Magazine, vol. 46, no. 6, Jun. 2008, pp. 128 –137.

[6] L. Battestilli, et al., "EnLIGHTened computing: An architecture for co-allocating network, compute, and other Grid resources for high-end applications," International Symposium on High Capacity Optical Networks and Enabling Technologies, Nov. 2007, pp. 1–8.

[7] A. Takefusa, et al., "G-Lambda: Coordination of a grid scheduler and lambda path service over GMPLS", Future Generation Computing Systems, vol. 22 No. 8, Oct. 2006, pp. 868-875.

[8] G. Roberts, T. Kudoh, I. Monga, J. Sobieski, and J. Vollbrecht, "Network Service Framework v1.0", Technical Report GFD 173, OGF NSI-WG, 2010.

[9] J. Sobieski, "GLIF 2011 Rio NSI PlugFest Guide and Interoperability Challenge", OGF NSI WG, 2011.

[10] G. Roberts, T. Kudoh, I. Monga, and J. Sobieski, "NSI Connection Service Protocol v1.1," GFD-173, OGF NSI-WG, 2011.

[11] R. Krzywania, et al., Network Service Interface: Gateway for Future Network Services, Terena Network Conference, Jun. 2012, pp. 1-15.

[12] I. W. Harbib, Q. Song, Z. Li, and N. S. V. Rao, 'Deployment of the GMPLS Control Plane for Grid Applicaitons in Experimental High-Performance Networks', *IEEE Communications Magazine*, Vol. 44, Issue 3, March 2006, pp. 65-73.

[13] N. Charbonneau, V. M. Vokkarane, C. Guok, and I. Monga, "Advance Reservation Frameworks in Hybrid IP-WDM Networks," IEEE Communication Magazine, vol. 49, Issue 5, May 2011, pp. 132 –139

[14] Handling Dynamiche Lichtpaden Manual, Version 0.2, Published by SURFnet, Nov. 2008.

[15] Grid Network Service-Web Services Interface (GNS-WSI), version 3, 2008, "https://www.g-lambda.net" [retrieved: Sep. 2013].

[16] Y. Cha, K. Lee, C. Kim, J. Kong, J. Moon, and H. Lim," Grid Network Management System Based on Hierarchical Information Model", Communications in Computer and Information Science, 1, Vol. 206, Part 4, Sep. 2011, pp. 249-258.

[17] R. Hughes-Jones, Y. Xin, G. Karmous-Edwards, and J. Strand, "Network Performance Monitoring, Fault Detection, Recovery, and Restoration", Grid Networks, Editors: F. Travostino, J. Mambretti and G. Karmous-Edwards, Wiley, pp. 253-275.

[18] K. Ogaki, M. Miyazawa, T. Otani, and H. Tanaka, "Prototype demonstration of integrating MPLS/GMPLS network operation and management system", OFC 2006, Mar. 2006, pp. 1-8.

[19] Z. Zhao, et al., "Planning data intensive workflows on inter-domain resources using the Network Service Interface (NSI)," 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, Nov. 2012, pp. 150-156.

 ISBN: 978-1-61208-305-6

TABLE II. COMPARISON OF ADVANCE RESERVATION FRAMEWORKS

Framework	Provisioning Layer	Network Resource Provisioning System	Grid Co-scheduling Capabilities	Protection management per VC	With NSI
DynamicKL	Layer 2 & 3	Integrated (MPLS/VLAN)	No	Yes	Yes
OSCARS	Layer 2 & 3	Integrated (MPLS/VLAN)	No	No	IDC/NSI
OpenDRAC	Layer 1 & 2	Integrated	No	No (protection only)	Yes
EnLIGHTened	Layer 1 & 2	Integrated (GMPLS based)	Yes	No	No
G-Lambda A/K	Layer 1 & 2	Integrated (GMPLS based)	Yes	No	Yes
PHOSPHORUS	Layer 1 & 2	ARGON, DRAC and UCLP	Yes	No	No
AutoBAHN	Layer 2 & 3	Integrated	No	No	IDC/NSI
DRAGON	Layer 1 & 2	VLSR (GMPLS based)	No	No	No

 ISBN: 978-1-61208-305-6

Towards Automating Mobile Cloud Computing Offloading Decisions: An Experimental Approach

Roberto Beraldi, Khalil Massri
Computer and System Science Department
La Sapienza University of Rome
Rome, Italy
{beraldi, massri}@dis.uniroma1.it

Abderrahmen Mtibaa, Hussein Alnuweiri
Department of Electrical and Computer Engineering
Texas A&M University
Doha, Qatar
{amtibaa, alnuweiri}@tamu.edu

***Abstract*—Mobile applications require more and more resources to be able to execute tasks on a single device, despite the fact that mobile devices are getting better capabilities. This has been addressed through several proposals for efficient computation offloading from mobile devices to remote cloud resources or closely located computing resources known as cloudlets. In this paper we adopt a experimental driven approach to highlight the offloading tradeoffs. We show that rather than always offloading tasks to a remote machine, running particular tasks locally can be more advantageous. We propose a novel generic architecture that can be integrated to any mobile cloud computing application in order to automate the offloading decision and help these applications to improve their response time while minimizing the overall energy consumed by the mobile device.**

***Index Terms*—Mobile Cloud Computing, Chess Game, Android Experimentation.**

I. Introduction

Mobile devices are increasingly utilized beyond simple connectivity for services that require more complex processing and capabilities. These include pattern recognition to aid in identifying snippets of audio or recognizing images, reality augmentation to enhance our daily lives, collaborative applications that enhance distributed decision making, planning and coordination. These applications are already in ubiquitous use today, others are still prototypes awaiting the next generational change in device capability and connectivity.

Cloud computing, in general, is reshaping the design and implementation of today's software applications. These applications are designed originally for desktops that are always connected to the Internet. While traditional cloud applications has been quite successful, to-date it suffers from a number of shortcomings especially with the presence of wireless communication at the edge. Shortcomings include the high latency and energy consumption caused by the intermittent aspect of wireless networks, which makes executing tasks locally more advantageous in certain circumstances.

In this paper, we adopt an experimental based approach in order to highlight the need of an automated offloading mechanism which decide whether a given task should run locally or remotely at the cloud. We propose a generic middleware architecture that can be plugged into any mobile cloud computing (MCC) application. For a specific task, based on the task characteristics and device capabilities, our architecture decides whether the task should be offloaded to a distant cloud or run locally on the device itself. We then implement a chess game as prototyping mobile application in order to identify under which circumstances would migrating be advantageous. We used the chess game as a real testbed environment to identify all the factors that help make an efficient offloading decision with respect to users' preferences or minimizing the overall resources usage.

The rest of this paper is organized as follows. Section 2 briefly discusses work related to our research. In Section 3 we describe the design of our generic architecture for mobile cloud computing applications. Section 4 describes our experimental platform. Section 5 presents the results from an experimental evaluation. Section 6 summarizes our findings and discusses our future work agenda in this area.

II. Related Work

Leveraging mobile networking and cloud computing attracts many researchers nowadays [4]. [6] was one of the earliest solutions for dynamic partitioning among mobile computers and a fixed infrastructure. There are a number of offloading frameworks that can support the development of offloadable applications. Offloading can be achieved at the level of services, methods and system. Service offloading intercepts those parts of the code that a software developer has manually set up for offloading. Cuckoo [5] integrates into Android applications by creating a proxy inside the application for the interfaces that the application developer has defined. The proxy then decides whether to invoke its corresponding local method or to migrate the computation to the surrogate. Method offloading, however, uses per-method annotations and wraps methods directly for proxying. This approach is less intrusive from application developer's viewpoint in the sense that it does not conceptually require strict separation of offloadable code parts. MAUI [3] implements this ideology. It investigates the energy consumption challenge when offloading computationally heavy tasks to a cloud rather than executing locally. MAUI relies on developer effort to convert mobile applications to support such decision making, and secondly, it only considers the possibility of offloading to different types of infrastructures. CloneCloud [2] presents a solution which decides whether to execute a task on a remote cloud

 ISBN: 978-1-61208-305-6

service versus executing it locally based on static analysis and dynamic profiling information of a task. CloneCloud, on the other hand, uses a modified virtual machine implementation of Android to intercept running threads at byte-code level and to migrate them for distributed concurrency. As a side effect in reducing burden to the application developer, image-level offloading frameworks are required to be more sophisticated.

In fact, with the advent of mobile device capabilities, migrating task computation always to powerful machine is questionable. We believe that running certain tasks locally on mobile devices can be more advantageous and may save both energy and time especially in the presence of intermittent wireless connectivity. In this paper, we adopt an experimental approach towards identifying the potential gain of mobile computational offloading in regards of both response time and energy consumption, and propose a design of a novel generic architecture that help actual application to make the best offloading decision based on these metrics.

III. Mobile Cloud Computing Architecture: Migrate vs. Run Locally

Mobile cloud computing is indeed becoming a dominant trend. However, current systems mainly focus on (i) offloading all functionalities to the cloud via simple client server architectures, or (ii) implementing applications and services that run locally on the mobile device. In this paper, we propose a novel architecture that leverages the two previous cases and computes, in runtime, the best offloading method with regards to two main metrics: the total response time and the overall energy consumed by the mobile device.

As summarized by Figure 1-(a), our architecture proposes a generic middle-ware that can be plugged into any mobile cloud computing application. Such architecture receives a given task T from the mobile computing application and based on the device capabilities (*e.g.,* CPU usage, memory, available energy), it computes an utility function which helps deciding whether the task T should be offloaded to a distant cloud or run locally on the device itself.

Task Modeling Engine: this engine is responsible of receiving tasks from the application. It models each task T by a combination of data, D_T, taken as input to perform such a task, and computation, C_T, that the task needs to perform on this data in order to yield a result. A given application consists of many tasks, and the more data and computation intensive these tasks are, the more time and energy required to perform them.

Decision Maker: it is the main engine of our architecture. It uses data from the device data base and cache and triggers the estimators engines in order to make a final decision about the offloading method of the given task T. It receives T from the task modeling engines and computes an utility function in order to send back the task to the application for local execution, or forward it the task forwarder for remote execution at the cloud.

Energy and Delay Estimators: the energy and delay estimators are responsible of estimating (i) the approximate energy to be consumed after running the task locally or remotely at the cloud, and (ii) the total response time $t_T = t_T^{end} - t_T^{beg}$, where t_T^{end} and t_T^{end} represent respectively the time in which the application receives the results of the task T, and the time in which the application sent the task to the task modeling engine. The estimators take as input, the task characteristics which are D_T, and C_T in addition to historical decisions of similar tasks (stored in the task/decision cache)

Task Forwarder: Upon receiving an order to migrate the task to distant cloud via the decision maker, the forwarder forwards the task to the distant cloud and keeps track about the status of the connection. In case of intermittent connectivity the forwarder reports an additional delay to the decision maker which will remake the decision based on the new factors (additional estimated delay).

Databases and Cache: the databases store the device capabilities, the communication technologies (*e.g.,* wireless, 3G, 4G, Bluetooth) and the task/decision cache. The cache is implemented to avoid remaking decisions for similar or identical tasks.

This architecture is built on the promise of computational offloading gain experienced, either in energy saving or task completion time, by any given cloud. If no potential gain/advantage exists, there is no point in adopting this architecture in the first place. Consequently, we believe that the major problem that needs to be addressed at this early stage of research is answering the question of when should we offload a given task. Therefore, we focus our attention in the remainder of this paper to quantitatively verify the potential gain tradeoff between energy and time while executing the task locally or offloaded remotely.

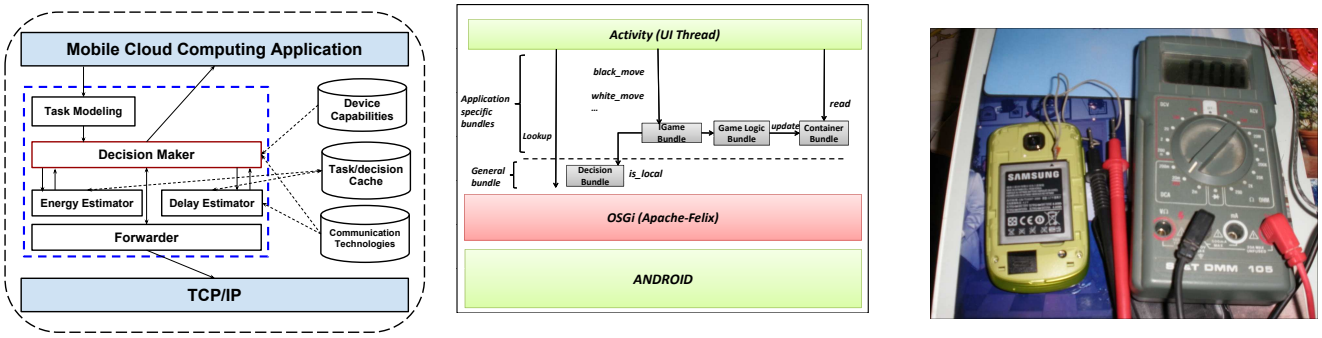
IV. Experimental Testbed: Chess Game Prototype

Our goal consists of making "good" decision about migrating computation of a given task or running it locally. It involves making a decision regarding which task is worth offloading. In order to answer this question, we adopt an experimental approach using a real testbed environment. Our goal is mainly to: (i) identify the trade-off between the gain of migrating tasks as opposed to running them locally, and (ii) identify under which circumstances would migrating be advantageous.

We choose to implement a chess game as it has a single task whose complexity can be easily set according to the expertise level, *e.g.,* from beginner to expert.

Our application is divided into three major software layers as shown in Figure 1-(b). The bottom layer is the standard OSGi's implementation of the Apache Software foundation community, Felix [1]. This software layer basically allows to register and lookup for other OSGi bundles. We also implement the following four bundles; (1) The *decision* bundle that encapsulates the decision logic about running an application module locally, i.e., in the device, or remotely in the cloud. It roughly corresponds to the Decision Maker component of Figure 1. (2) The `IGame` bundle is the interface seen by the

 ISBN: 978-1-61208-305-6



(a) Our high level architecture (b) Prototyping implementation (c) Energy testbed

Fig. 1. System architecture and experimental testbed

android application. This bundle implements the following operations: initBoard, initSearch, blackMove and whiteMove. (3) The `GameLogic` bundle implements the computer move and occupies 456 KB. This is the bundle (Task) to be uploaded into the cloud for implementing remote calls. (4) The `container` bundle is used as access point to read the data generated after a move.

A game tree for the chess game is used to decide the best counter move. It is composed of nodes representing the state of the board and edges the possible moves. The tree is explored at different levels, according to the difficulty degree, labeled from 1 (beginner) to 4 (expert). The maximum explored level of the tree was, respectively, 3, 7, 10, 14. The best counter move is searched applying the alpha-beta searching algorithm. This algorithm exploits lower and upper bounds (named alpha and beta), to prune part of the game tree that cannot possibly influence the final decision.

As far as the cloud technology is concerned, we have used OpenShift, a Red Hat's free, auto-scaling Platform as a Service (PaaS) for applications. It allows to run different VM, called gears. It uses the notion of cartridge, a set of predefined technologies that can be installed and run inside gears. In addition, Openshift allows users to exploit the powerful Do-It-Yourself (DIY) feature that allows for running just about any program that speaks HTTP. As the Apache felix is not available as a cartridge we have used the DIY feature for our experiment. Moreover, in order to allow android openshift communication we have exploited Apache CXF, an open source services framework that helps to build and develop services using frontend programming APIs, like JAX-WS and JAX-RS. In particular we have exported the service in the cloud as RESTful HTTP endpoints . CXF allows to export OSGi services as Web Services.

V. Experimental Analysis

We have tested the application in three different settings, named local, cloud and cloudlet. In the local mode, the mobile device executes all the code, whereas in the cloud and cloudlet modes, the device only executes the code required for UI updating and the user moves. In the cloudlet node, the remote task is executed in a VM hosted by a server machine located in the same wireless LAN of the device.

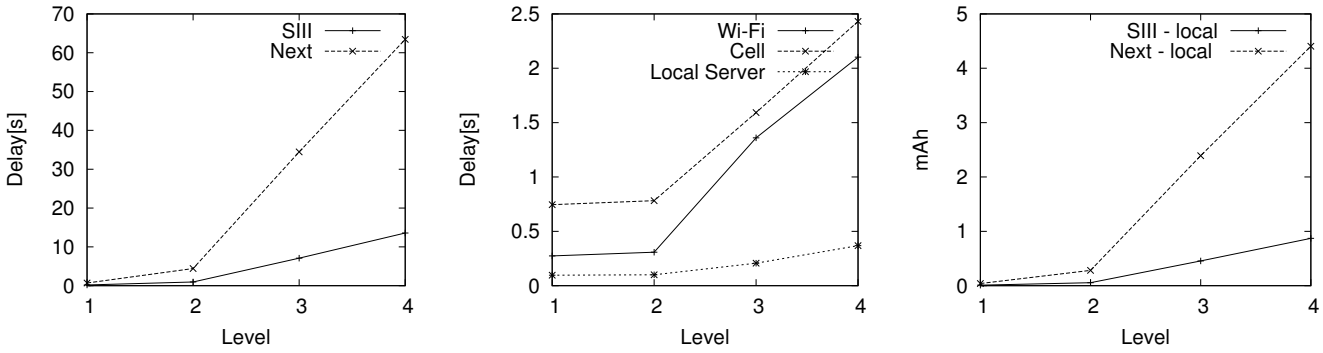
The opening phase in a chess game is a very important phase, as it may shape the way the whole match will proceed. The reply to an opening is then a good situation game to test the performance of our application. We have considered a famous openings due to Anderssen, A2 in A3. Two different mobile devices have been used in the experiments; Samsung Galaxy Next Turbo and a Galaxy SIII. In the cloud mode, the game logic runs in the openshift platform and Internet is accessed either through a wi-fi or via GSM. Finally, in the cloudlet setting, the same public cloud environment runs locally on a PC running Windows 7 Home Premium, CPU i7-3610QM CPU @2.30GHz with 8 GB RAM.

Response Time: The time elapsed from when the white ends to move till the black move ends is called the response time. To measure such a response time the same opening was repeated three times. Figure 2 shows the average response time for (a) local execution and (b) remote execution. Under the local execution case, it is clear that as the complexity of the task increases, as in level 3 and 4, the response time increases considerably, especially, for the lower computation capability one (Next). While, on the other hand, when the task execution is done remotely, either on the cloud or cloudlet, the response time is always acceptable, even when the task complexity is the maximum. The level can thus be used both to describe the task complexity and to drive decision maker to select the execute mode. In addition it also affects the power consumption, as we discuss in the following subsection.

Energy Consumed Locally: Energy saving is one of the most important expected benefit that mobile cloud computing should provide. Figure 2-c reports the average mAh for the local execution mode. This plot is similar to the delay's one. This is due to the fact that the delay to the reply is indeed due to the computation of the black move. In other words, if the computation requires T s then the required charge is kT, where k is a constant independent from T.

Quantifying the Overhead: In the previous section, we were assuming that the cloud implements already chess game algorithms and maintains a state about each player of a game. In this section, we investigate a scenario where the application is totally implemented in the phone device itself and a remote execution of a task requires transferring a chunk of data that is needed to run the tasks remotely.

 ISBN: 978-1-61208-305-6



(a) Response time for local execution (b) Response time for remote execution (c) Energy vs. difficulty levels

Fig. 2. Response time and energy performance of local and remote execution

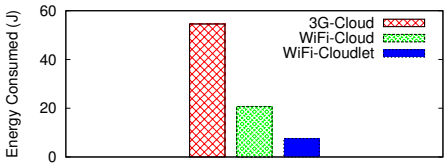


Fig. 3. Energy consumed while offloading to cloud or cloudlet using SIII

We have measured the number of bytes exchanged between the client and cloud during the initialization phase (when bundles are emitted to the cloud) and during moves. For this purpose we have used the wire shark sniffer. We then run a set of experiments to quantify the energy overhead related to communication between the mobile device and the cloud as shown in Figure 1-(c). We remove the battery of the Samsung SIII device and solder wires coming from a power supply into the battery contacts of the device. The power supply that we use comes with a built in ammeter and voltmeter. We then provide a constant voltage according to the manufacturer specifications and power the device on. Using the current and voltage readings from the ammeter and voltmeter respectively, we are able to determine the power being consumed by the phone at any instance.

We run each experiment 5 times and compute the energy readings while idle, and those while making wireless transfers (sending or receiving data). Figure 3 compares the energy consumed while transmitting the initial code to a distant cloud or a nearby cloudlet. We consider 3G and WiFi as two wireless technologies to communicate with a distant cloud and WiFi to communicate with cloudlet.

Figure 3 shows that 3G consumes more than double the energy required to communicate using WiFi. This may imply running tasks locally if only 3G connectivity is available to reach a distant cloud. Using WiFi, communicating with a nearby cloudlet is 2 to 3 times less expensive than communicating to a cloud. This confirms the results showing in [7]. However, the cloudlets are less computationally powerful than a cloud which may introduce additional delay and therefore energy to finish executing the task.

VI. Summary & Discussion

We believe that there is a need for a generic and flexible framework that can be integrated by any mobile cloud computing to automatically decide based on a given characteristic of the application or its tasks whether it is better to run execution locally or remotely. We begin by running a set on preliminary experiments in order to identify a tradeoff between two conflicting metrics. Our preliminary experimental results have shown that energy and time depend on the computation complexity of the task and the device capabilities. On the other hand, when offloading the task to the cloud the energy and time consumed will highly depends on the communication technology used. In our work, we have adopted a first step towards automating an offloading decision maker, which measures the task complexity, then based on the device capability and the communication heuristic profile, it can choose the better choice. We plan to expand our experimental and analytical results to help identify the objection function that will be used by our decision maker. The objection function main goal is to quantify the gain based on user preferences or simply minimizing the overall resource usage.

Acknowledgment

This research was supported by a research grant from the Qatar National Research Fund under project NPRP 09-1116-1172.

References

[1] Apache felix. http://felix.apache.org/.

[2] B.-G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti. Clonecloud: elastic execution between mobile device and cloud. In *Proceedings of the sixth conference on Computer systems*, EuroSys '11, pages 301–314, New York, NY, USA, 2011. ACM.

[3] E. Cuervo, A. Balasubramanian, D. ki Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl. Maui: making smartphones last longer with code offload. In *MobiSys'10*, pages 49–62, 2010.

[4] J. Flinn. Cyber foraging: Bridging mobile and cloud computing. *Synthesis Lectures on Mobile and Pervasive Computing*, 7(2):1–103, 2012.

[5] R. Kemp, N. Palmer, T. Kielmann, and H. Bal. Cuckoo: A computation offloading framework for smartphones. pages 59–79. Springer Berlin Heidelberg, 2012.

[6] A. Rudenko, P. Reiher, G. J. Popek, and G. H. Kuenning. Saving portable computer battery power through remote process execution. *SIGMOBILE Mob. Comput. Commun. Rev.*, 2(1):19–26, Jan. 1998.

[7] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies. The case for vm-based cloudlets in mobile computing. *Pervasive Computing, IEEE*, 8(4):14–23, 2009.

 ISBN: 978-1-61208-305-6

3D On-chip Data Center Networks Using Circuit Switches and Packet Switches

Takahide Ikeda Yuichi Ohsita, and Masayuki Murata
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
{t-ikeda, y-ohsita, murata}@ist.osaka-u.ac.jp

***Abstract*—The energy consumption of the data center becomes a great problem. One approach to reduce the energy consumption of the data center is to use *on-chip data centers*, which are integrated circuit chips that perform the tasks in a data center. On-chip data centers are constructed of cores and the network between cores. Because the tasks in the data center are performed by the cooperation between servers, the network between cores in the on-chip data center may have a large impact on the performance of the chip. In this paper, we investigate the network structures for the on-chip data centers. We focus on the 3Dnetwork using both circuit and packet switches, and compare the energy consumption and the delay of the candidate network structures. The results show that (1) the servers should connect to the packet switches in the same layer, (2) the packet switches should connect to the circuit switches in all layers, and (3) the layer including both of circuit switches and packet switches should be avoided to reduce the energy consumption and the delay.**

***Keywords*—network on chip; data center; energy consumption; delay; 3D on-chip network**

I. INTRODUCTION

In recent years, online services such as cloud Computing have become popular, and the amount of data, required to be processed by such online services is increasing. Such a large amount of data is handed by data centers, and many data centers have been built. As the services provided by data centers become popular, the energy consumption of the data center becomes an important problem. The energy consumed by data centers occupies 1.5% of the total energy consumption consumed in the world [1]. Thus, an energy efficient data center is required.

A data center is constructed of many servers. In a data center, each server performs its assigned task, cooperating with other servers [2]. The data center can process a large amount of data because a server cooperates.

One approach to reduce the energy consumption caused by the data center is to make an integrated circuit chip that can perform the above tasks in a data center. This kind of chip is called an *on-chip data center* [3]. An on-chip data center is made of a large number of CPU cores and the network between the cores on a single chip. An on-chip data center works with a significantly small energy because of its small wiring length of the network within a chip [4].

Most of existing work on on-chip data centers focus on the usage of many cores on the chip. However, because tasks in a data center require communication between servers, the network structures between cores may have a large impact on the performance and/or the energy consumption of the on-chip data center.

The network within a chip is often called a *Network on chip (NoC)*, and constructed of switches. Two types of switches are used in a NoC: packet switches and circuit switches.

A packet switch relays packets or flits, which are a small pieces of a packet, based on their destination addresses. On the other hand, a circuit switch connects its input port with one of its output ports based on the configuration. A circuit switch consumes a small energy compared with a packet switch because it does not require any processing to relay traffic, though multiple flows from different input ports cannot share the same output port.

Several NoC architectures that use both packet and circuit switches have been proposed [5-7]. In these architectures, the circuit path between packet switches is established by configuring the circuit switches along the route of the circuit path. The set of the packet switches and the established circuit paths constructs the network topology. In these architectures, the network topology can be changed by the configuration of the circuit switches. Stensgaard et al. [7] proposed a method to configure the circuit switches suitable to the application before starting the application.

The network architectures using both of packet and circuit switches are also effective in an on-chip data center. In a data center, though the traffic pattern changes significantly and frequently, each server communicate with only a small number of servers at once [8]. Considering such traffic, the network topology where the communicating server pairs are connected closely is preferable. This network topology can be set by setting the circuit switches in the network using both of the packet and circuit switches. Even if the traffic pattern changes, we change the network topology so as to suit the current traffic pattern by reconfiguring the circuit switches.

In recent years, another new NoC architecture called *3D NoC* has been proposed [9]. The 3D NoC is constructed by stacking multiple 2D chip layers vertically. The vertically stacked layers decrease the number of hops between switches. Moreover, the vertical links of the 3D NoC are significantly shorter than the horizontal links. As a result, the 3D NoC reduces both the energy consumption and the delay.

In addition, the 3D NoC improves the effectiveness of using packet and circuit switches. Because the 3D NoC increases the number of candidate routes of the circuit paths, we establish more circuit paths, which reduce both the energy consumption and the delay. However, the 3D NoC using both packet and circuit switches has not been discussed sufficiently.

In this paper, we investigate the network structures suitable for the on-chip data center. In an on-chip data center, a server is constructed by multiple directly connected cores. Then,

 ISBN: 978-1-61208-305-6

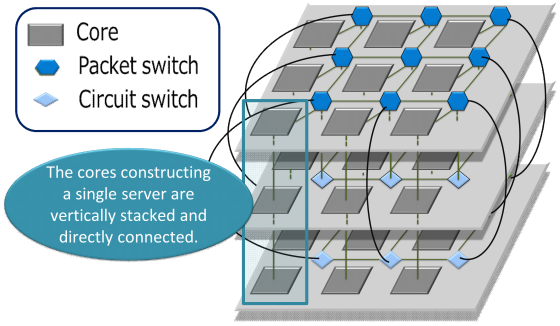


Figure. 1. 3D on-chip data center network

the network connects the servers. In this paper, the network between servers is constructed as a 3D network using circuit and packet switches. We investigate the network structures, focusing on the following three points; (1) connection between layers in the 3D network, (2) connection between servers and switches, and (3) placement of switches within each layer. The results show that (1) all servers should be connected to the packet switches in the same layer, (2) all packet switches should be connected to all layers, and (3) each layer should have only the same type of switches.

The rest of this paper is organized as follows. Section II explains the overview of the on-chip data center used in this paper. In Section III, we investigate the network structures suitable to the on-chip data center. Section IV presents the conclusion.

II. On Chip Data Center Network

A data center is constructed of servers and a network between servers. The tasks in a data center, such as handling a large amount of data, are performed by servers cooperating with each other. Such a task in a data center is split into subtasks, and each subtask is assigned to and performed by one of the servers. Each server obtains the data or the results of the other subtasks from the other servers, if the data or the results are required to complete its subtask.

In this paper, we investigate the on-chip data center, which performs tasks in a data center. The on-chip data center used in this paper is constructed of multiple cores and a network between cores. The tasks are handled by multiple cores in a data center. Each of the cores in the on-chip data center provides a resource or cache memory. The related multiple cores are connected to each other, and act as a single server in a data center. We call these connected multiple cores a *server* in the on-chip data center. The network structure used in this paper is shown in Fig 1. In this structure, the cores constructing a single server are vertically stacked and directly connected. Then, the servers are placed in a lattice.

The network between servers is constructed of switches placed in a 3D lattice, because the lattice network can be easily constructed on a chip. Each server is connected to the network by connecting one of its core to one of the switches.

In the on-chip data center, we use two kinds switches, i.e., packet switches and circuit switches. The packet switches and the circuit switches have their advantages and disadvantages. The circuit switches consume less energy than the packet switches, because the circuit switches do not require complicated processing such as decision of the next hop. However, the circuit switch cannot relay flows from different input ports to the same output port. On the other hand, the flows from the different input ports share the same output port in the packet switch, though the packet switch consumes more energy.

In this paper, we use both types of switches as follows. All servers are connected to packet switches, so that each server can communicate with multiple servers at once. The switches not connected to servers are circuit switches because the circuit switches consumes less energy. In this network, the traffic is sent after constructing the network topology by setting the circuit paths between packet switches. The circuit paths are established by configuring the circuit switches along the paths. Then, the traffic is sent over the network topology of the packet switches constructed by the circuit paths.

This network structure has the following parameters; (1) the connection between layers, (2) the layers where switches connected to servers are deployed, and (3) the types of switches deployed in each layer, which are discussed in Section III.

III. Comparison of on-Chip Data Center Networks

A. Network structures

In this section, we investigate the following parameters of the 3D network structures for on-chip data center.

1) Inter-Layer Connection: There are two types of the inter-layer connection. The first type is shown in Fig 2(a). In this case, switches in all layers are connected to the same packet switch. We call this type of connection the *packet switch centric connection*.

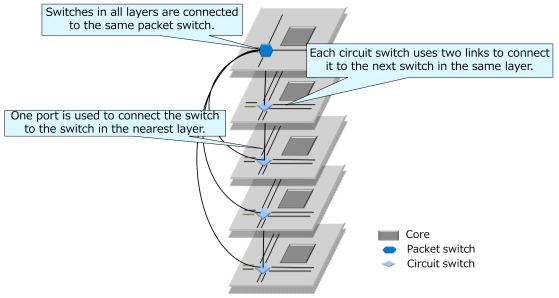
Another type of the inter-layer connection is shown in Fig 2(b). In this case, all vertical links are constructed only between nearest layers. We call this type of connection the *nearest layer connection*.

In our comparison described in Subsection III-B1, we deploy all packet switches at the first layer. All packet switches have 8 ports and all circuit switches have 10 ports in both types of connections. We set the number of layers to 5. In the circuit switch centric connection, each circuit switch uses two links to connect it to the next switch in the same layer. That is, each circuit switch uses 8 ports to the connection in the same layer. Two ports per circuit switch are used to the vertical connection. One of them is used to connect the switch to the packet switch at the first layer. The other port is used to connect the switch to the switch in the nearest layer.

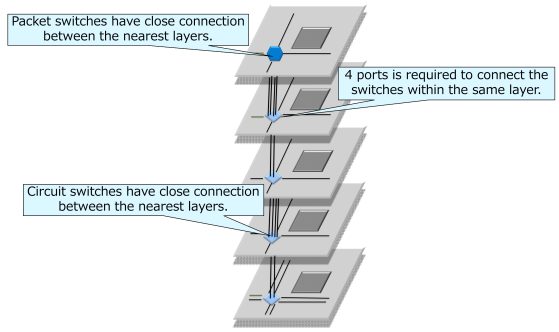
In the nearest layer connection, we add close connection between the nearest layers. All vertical links from the packet switches at the first layer are connected to the switches at the second layer. Thus, 4 ports of the switches at the second layer are connected to the switches at the first layer. Among the residual ports, 4 ports are required to connect the switches within the same layer. Finally, the other ports are used to connect the switches to the third layer. The switches in the other layers are connected in a similar way.

 ISBN: 978-1-61208-305-6

The comparison of these connections clarifies whether the one hop connection from the packet switches to any layers is preferable or the close connection between the nearest layer is preferable.



(a) Packet switch centric connection



(b) Nearest layer connection

Figure. 2. Inter layer connection

2) Layer of Swithes Connected to Servers: In the on-chip data center investigated in this paper, each server is connected to one of the switches nearest to the server. As shown in Fig 3, there are two types of connections between servers and switches. In the first type of connection, all servers are connected to the switches in the same layer. We call this type of connection the *same layer connection*. In the other type of connection, the servers neighboring with each other are connected to the switches in the different layers.

In the same layer connection, the number of hops between servers is small because all servers are connected in the same layer. However, the connections of packet switches at the first layer are static. On the other hand, the connections between

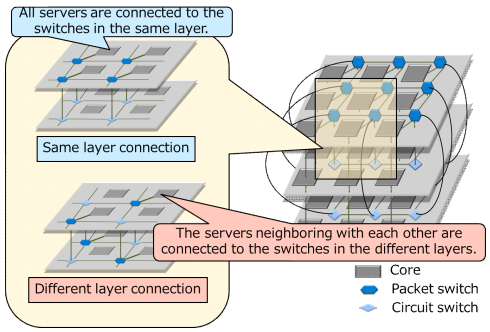
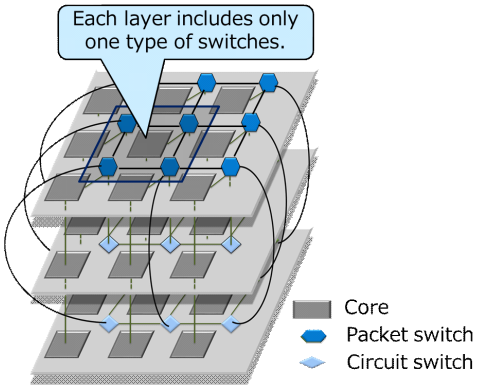
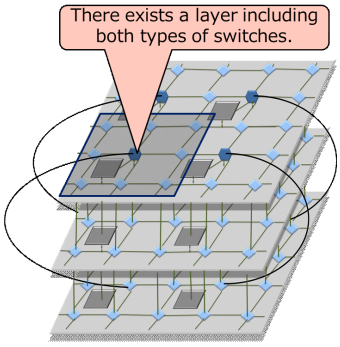


Figure. 3. Connection from servers



(a) Single type placement



(b) Multiple type placement

Figure. 4. Placement of switches within each layer

packet switches can be changed in any layers in the different layer connection.

3) Placement of Switches within a Layer: There are two kinds of placement of the switches in the same layer. The first one is shown in Fig 4.(a). In this type of the placement, each layer includes only one type of switches. We call this type of the placement the *single type placement*. In the other type of placement, there exists a layer including both types of switches. We call this type of placement the *multiple type placement*.

The *multiple type placement* has more candidates of routes of circuit paths between the packet switches than the *single type placement*. Thus, the energy efficient routes may be found, even when the number of flows to be accommodated is large. However, the number of switches passed by each flow between servers increases. On the other hand, the *single type placement* has less routes between the server pairs. However, the numbers of switches passed by each flow between servers are small.

B. Models Used in Our Comparison

1) Energy consumption model: The energy consumed by the network on chip depends on (1) network structure, (2) the traffic amount on the network, and (3) the bit flips of the traffic.

Wolkotte et al. [10] model the energy consumed by a circuit switch, a packet switch and a link in the NoC, in the case of 50% bit flips. In this model, the circuit switch consumes 0.37 μW, the packet switch consumes 0.98 μW, and the link consumes $(0.39 + 0.12L)$ μW where L is a length of link

 ISBN: 978-1-61208-305-6

(mm) to relay 1 bit of traffic. In this paper, we use this model to evaluate the energy consumption. In this paper, we focus only on the energy consumed by the network, and exclude the energy consumed by the cores.

Though the actual energy consumed by the switches and links may be different from this model, the packet switch consumes more energy than the circuit switch in any cases because the packet switch requires more process such as checking the destination of each packet. Therefore, the results in this paper are independent of the switch architectures.

2) Delay Model: In this paper, we also compare the delay between cores. We define the delay as the time required to receive all traffic by the destination cores after generating the traffic demands.

In the network on chip, packets are generally divided into flits, and each switch relays the flits. In this paper, we assume that each flit can be relayed by a packet switch to the next packet switch in 1 clock cycle. Though, the clock cycle required to relay a flit depends on the switch architectures and may be different from this model. The suitable network structures discussed in this paper are independent of switch architectures because the order of delays is the same as the results in this paper even if multiple clock cycles are required to relay a flit.

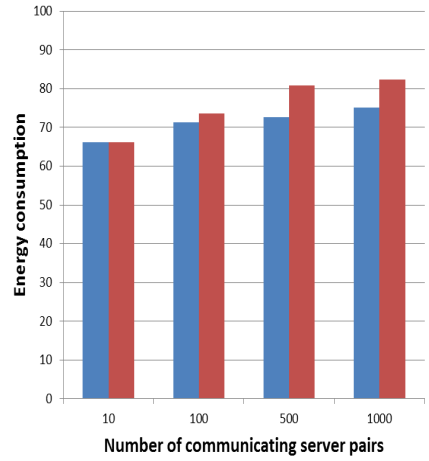
In the on-chip data center, we also use the circuit switches. The circuit switch is configured to connect the input and output ports in advance. The packet switches can be connected by configuring the circuit switches. The packet switch pairs, connected by the circuit paths, relay the flits by the same way as the packet switches which are directly connected to each other. The relay of the flits by the circuit switch takes no clock cycles. Thus, the delay between cores depends only on the number of packet switches passed by the flow.

3) Traffic Model: According to Benson et al. [8], each server communicates with only a small number of servers at once. Therefore, we generate traffic between randomly selected server pairs. In our evaluation, we vary the number of communicating server pairs from 10 to 1000. The traffic rates between selected server pairs are set to 10000.

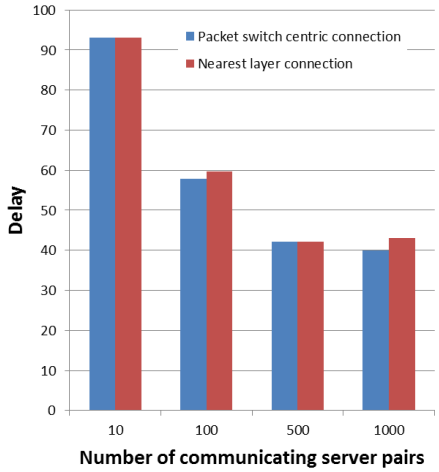
4) Path Computation Model: We calculate the routes of traffic so as to make the energy consumed by the traffic small.

In this paper, we calculate the route of all traffic demands. The route of each traffic demand is calculated by the Dijkstra algorithm setting the weights of the links to the energy consumed to relay the traffic. If the calculated route uses the circuit switch, we connect both ends of the input and the output ports, and remove the ports of the circuit switch before the calculation of the routes of the next traffic demands, so as to avoid the output ports of the circuit switch used by the other traffic from the different input ports.

In this path computation, we assume that the traffic demands are known before calculating the routes. By using this model, we discuss the suitable network structure when the routes are calculated optimally. However, the actual traffic demands may be unknown when calculating routes, and we require a method to calculate the routes without traffic demand information, which is one of our future work.



(a) energy consumption



(b) Delay

Figure. 5. Comparison of inter-layer connections

C. Results

In our evaluation, we use the network structure with 5 layers and 255 servers. We generate 4 patterns of traffic, and compare the average of the energy consumption and the delay.

1) Comparison of Inter-layer connections: In this subsection, we compare the network structures of different inter-later connections. The comparison of energy consumption is shown in Fig 5(a). The vertical axis of the figure indicates the energy consumption normalized so that the energy consumption in the 2D lattice using only the same number of packet switches as the 3D network structures used in this comparison becomes 100. Fig 5(a) shows that both of the inter-layer connections reduce the energy consumption compared with the 2D lattice. This is because the 3D lattice structures used in this comparison establish the circuit paths to reduce the energy consumption. However, the energy consumption in both types of the connections becomes close to that of the 2D lattice when the number of communicating server pair becomes large. This is because we cannot establish energy efficient circuit paths for all communicating server pairs. As a result, a large amount of traffic passes multiple packet switches similar to the 2D lattice.

 ISBN: 978-1-61208-305-6

Fig 5(a) also shows that the energy consumption of the packet switch centric connection is smaller than the nearest layer connection. This is because a flow is required to pass multiple layers to use the circuit switch whose layer is far from the packet switch in the nearest layer connection. Because each switch relaying the traffic consumes energy, the large number of switches passed by each flow cause a large energy consumption. On the other hand, the packet switches are directly connected to all layers in the packet switch centric connection. Thus, the number of switches passed by traffic is smaller than the nearest layer connection. As a result, the packet switch centric connection accommodates traffic with a smaller energy consumption than the nearest layer connection.

The comparison of delay is shown in Fig 5(b). The vertical axis of the figure indicates the delay normalized so that the delay in the 2D lattice using only packet switches becomes 100. Fig 5(b) shows that the 3D network structures using both of circuit switches and packet switches reduce delay significantly compared with the 2D lattice. This is because the circuit paths reduce the number of hops of packet switches.

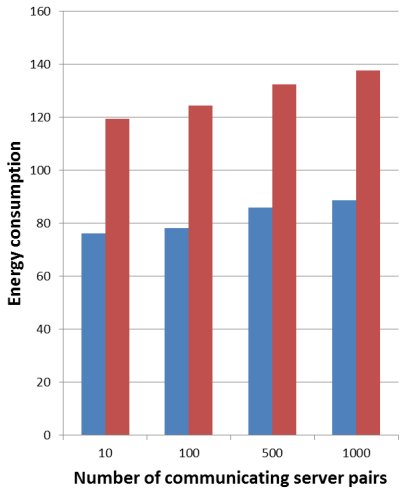
The normalized delay becomes small when the number of communicating server pairs increases. This is because the circuit paths balance the loads. In the case of the 2D lattice, traffic concentrates at some packet switches, and is required to wait to be relayed, when the number of communicating server pairs becomes large. On the other hand, in the 3D lattice using circuit switches and packet switches, the circuit paths directly connect the packet switches which are far from each other, and avoid concentration of traffic on a certain switch.

Fig 5(b) shows that the packet switch centric connection and the nearest layer connection achieve the similar delay. This is because the circuit switches do not have an impact on the delay though the traffic passes more circuit switches in the nearest layer connection than the packet switch centric connection.

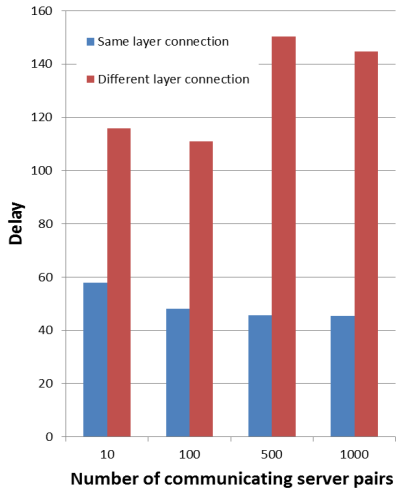
2) Comparison of the Connection between Servers and Switches: In this subsection, we compare the network structures of the different types of the connections between servers and switches.

The comparison of energy consumption is shown in Fig 6(a). The vertical axis of the figure indicates the energy consumption normalized so that the energy consumption in the 2D lattice using only packet switches becomes 100. Fig 6(a) shows that the same layer connection achieves the lower energy consumption than the 2D lattice. On the other hand, the different layer connection consumes more energy than the 2D lattice. This is because the traffic between servers passes more switches in the different layer connection than the 2D lattice.

The comparison of delay is shown in Fig 6(b). The vertical axis of the figure indicates the delay normalized so that the delay in the 2D lattice using only packet switches becomes 100. Fig 6(b) shows that the delay in the different layer connection becomes larger than the 2D lattice. This is because of the large number of hops of packet switches. In the different layer connection, the packet switches exist not only at the first layer, but also the other layers. Such packet switches block the long circuit path, and even cause a large number of packet switches passed by a flow. On the other hand, the same layer



(a) Energy consumption



(b) Delay

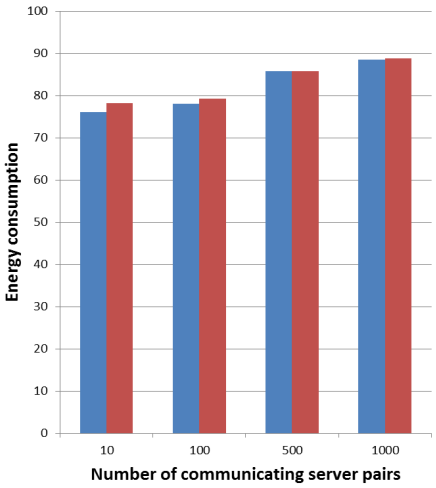
Figure. 6. Comparison of the connection from servers

connection reduces the delay significantly compared with the 2D lattice. This is because long circuit paths are established in the same layer connection, and reduce the number of packet switches passed by a flow.

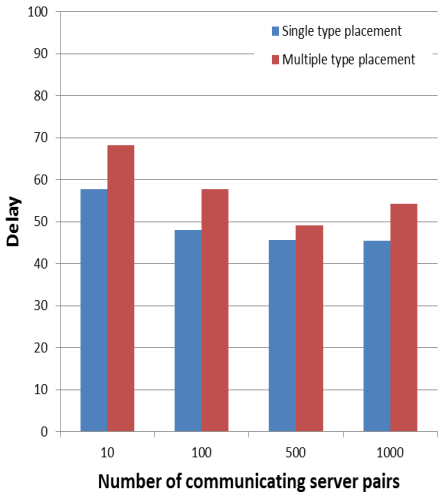
Similar to the results in Fig 5(b), Fig 6(b) also indicates that the normalized delay becomes small in the same layer connection when the number of communicating server pairs becomes large. This is because the circuit paths balance the loads. On the other hand, the normalized delay becomes large in the different layer connection. In the different layer connection, we cannot add the long circuit paths. Moreover, the number of packet switches passed by a flow is large. As a result, traffic concentrates at some packet switches.

3) Comparison of Placement of Switches within a Layer: In this subsection, we compare the types of the switches used in each layer. The multiple type placement increases the number of candidate routes for the circuit paths. However, the number of hops becomes larger than the single type connection. Comparing them, we clarify whether the larger

 ISBN: 978-1-61208-305-6



(a) Energy consumption



(b) Delay

Figure. 7. Comparison of the placement of switches within each layer

number of candidate circuit paths is preferable or the smaller number of hops between servers is preferable.

The comparison of energy consumption is shown in Fig 7(a). The vertical axis of the figure indicates the energy consumption normalized so that the energy consumption in the 2D lattice using only packet switches becomes 100. Fig 7(a) shows that the single type placement and the multiple type placement consume the similar energy. The multiple type placement has more routes of the circuit paths between servers than the single type placement, and can find energy efficient routes. However, the number of hops between servers becomes large, which consumes more energy. In the case of our simulation, the amount of the energy reduced by using circuit paths equals the amount of the energy increased by the increase of switches relaying the traffic.

Similar to Fig 5(a), Fig 7(a) also shows that both types of the placements consume the similar energy to the 2D lattice using only packet switches when the number of communicating servers becomes large. This is because we cannot set the circuit paths for all communicating server pairs.

The comparison of the delays is shown in Fig 7(b). The vertical axis of the figure indicates the delay normalized so that the delay in the 2D lattice using only packet switches becomes 100. In Fig 7(b), the single type placement achieves the smaller delay than the multiple type placement. This is because the routes are set so as to make the energy consumption small. Though the multiple type placement has more candidates routes for the circuit paths, our route calculation selects the routes that make the energy consumption small, even if the routes cause the concentration of traffic.

In addition, Fig. 7(b) also indicates that the normalized delay becomes small when the number of communicating server pairs becomes large similar to Fig. 5(b) and Fig. 6(b).

IV. Conclusion and Future Work

In this paper, we evaluated the 3D on-chip network structures for the on-chip data centers, which uses both of the circuit and packet switches. According to the results, to reduce the energy consumption and delay, (1) the servers should connect to the packet switches in the same layer, (2) the packet switches should connect to the circuit switches in all layers, and (3) the layer including both of circuit switches and packet switches should be avoided.

Our future work includes the method to calculate the routes suitable to the on-chip networks.

Acknowledgments

This work was partially supported by JSPS Grant-in-Aid for Young Scientists (B)23700077.

References

[1] J. G. Koomey and P. D, "Growth in data center electricity use 2005 to 2010," *The New York Times*, Aug. 2011.

[2] D. Abts and B. Felderman, "A guided tour of data-center networking," in *Communications of the ACM*, vol. 10, Jun. 2012, pp. 44–51.

[3] M. Kas, "Toward on-chip datacenters: a perspective on general trends and on-chip particulars," in *The Journal of Supercomputing*, vol. 62, Oct. 2012, pp. 214–226.

[4] R. Iyer, R. Illikkal, L. Zhao, S. Makineni, D. Newell, J. Moses, and P. Apparao, "Datacenter-on-chip architectures: Tera-scale opportunities and challenges in Intel's manufacturing environment," in *Intel Technology Journal*, vol. 11, Aug. 2007, pp. 227–237.

[5] T. Bjerregaard and S. Mahadevan, "A survey of research and practice of network-on-chip," vol. 1-51, Mar. 2006.

[6] M. B. Stensgaard and J. Sparso, "ReNoC: A network-on-chip architecture with reconfigurable topology," in *Proceedings of the Second ACM/IEEE International Symposium on Networks-on-Chip*, Apr. 2008, pp. 55–64.

[7] M. Modarressi, H. Sarbazi-Azad, and M. Arjomand, "A hybrid packet-circuit switched on-chip network based on sdm," in *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition, 2009. DATE '09*, Apr. 2009, pp. 566–569.

[8] T. Benson, A. Anand, A. Akella, and M. Zhang, "MicroTE : Fine grained traffic engineering for data centers," in *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*, Dec. 2011, pp. 1–12.

[9] F. Li, C. Nicopoulos, T. Richardson, and Y. Xie, "Design and management of 3D multiprocessors using network-in-memory," in *Proceedings of ISCA*, Jun. 2006, pp. 130–141.

[10] P. T. Wolkotte, G. J. M. Smit, N. Kavaldjiev, J. E. Becker, and J. Becker, "Energy model of networks-on-chip and a bus," in *Proceedings of IEEE International Symposium on System-on-Chip*, Nov. 2005, pp. 82–85.

 ISBN: 978-1-61208-305-6

Architecture for Platform- and Hardware-independent Mesh Networks

How to unify the channels

Sebastian Damm, Michael Rahier, Thomas Ritz, Thomas Schäfer
Mobile Media and Communication Lab, FH Aachen
Aachen, Germany
s.damm@fh-aachen.de
rahier@fh-aachen.de
ritz@fh-aachen.de
thomas.schaefer@alumni.fh-aachen.de

***Abstract*—This paper will prove that mesh networks among different platforms and hardware channels can help to channel valuable information even if public telecommunication infrastructure is not available due to arbitrary reasons. Therefore, results of a simulation for mesh networks on mass events will be provided, followed by the developed architecture and an outlook on future research. The developed architecture is currently being implemented and field tested on mass events.**

Keywords-mesh networks; platform independence; mobile software.

I. INTRODUCTION

On mass events like music festivals, the cellular reception is often insufficient because Global System for Mobile Communications (GSM) cells tend to be overloaded. Especially in terms of security, this is a serious issue, as people might not be able to communicate with rescue forces. This can lead to catastrophes like the mass panic on the German Love Parade in 2010 where 21 people died. During investigation of this event, crowd scientist G. K. Still pointed out that missing communication was one of the key factors for what happened [1].

Nowadays, we are dealing with plenty of mobile hard- and software platforms like iOS, Android, Windows Phone and others. These platforms use all different kinds of connection channels such as Bluetooth, WiFi(-Direct), NFC, etc. That means that there are several possibilities to compensate the mentioned lack of connectivity. Mesh networks can be a solution where people keep connected on such mass events, without having any connection to a cellular network. With support of some well-placed infrastructure like access points, relevant data could be pushed into the crowd and then be routed or broadcasted to other persons from device to device (Fig. 1).

Figure 1: Illustration of information flow in the crowd

Unfortunately, there is no or just few integration of the different communication channels, even within single platforms, and there are even more difficulties when trying to interconnect different platforms. To negotiate these obstacles, this paper presents a platform and hardware-independent architecture that integrates all different network types into an abstract layer. This architecture is currently being implemented as a java base library, which is used on Android. There is also an implementation for Windows Phone. Further research on the possibility of an iOS implementation is ongoing. The main goal is to enable automatic interconnectivity between different mobile platforms without the need of user interaction and regardless to the used communication technology. This will provide better ways for promoters of mass events to reach their guests in case of emergency.

Following to this introduction, the paper will show related work in the field of mesh networks. In section three, a simulation will be presented, showing the possibility of creating a mesh network in the scenario of a mass event. After the general possibility is proven, an architecture for hardware- and software-independent mesh networks will be introduced in section four, followed by a conclusion and outlook to future research.

II. RELATED WORK

The research project “iWave” (information waves on mass events) addresses the problem of missing or poor connectivity on mass events. No connection means no ability to communicate in security related issues like mass panic, severe weather or just injuries. The architecture proposed in this article is part of ongoing research, where a reusable communicator component to span mesh networks will be implemented on different mobile platforms.

One project providing similar functionality is the middleware “Beddernet” [2]. It is capable of spanning mesh networks using Bluetooth. Unfortunately, there are several downsides of Beddernet. First, it is limited to Android and not available for other platforms. Second, it is using Bluetooth as the only channel, leaving out WiFi and such. The third problem is that the last change to the project was committed in August 2012 (and before that in July 2010); so, it can be assumed the project is not maintained anymore.

 ISBN: 978-1-61208-305-6

Another project dealing with the mentioned issues is the MANET project [3]. It uses sensor nodes based on the ZigBee standard (based on IEEE 802.15.4) [4]. A huge disadvantage is the use of special sensor nodes instead of the built in hardware in smartphones that people already carry around.

The University of Darmstadt proposed an approach to use WiFi routers in emergency cases to span ad-hoc networks [5]. This is suitable for communication within cities, but not transferrable to mass events, as the area is quite limited but crowded with lots of people with not as many routers as in an urban environment.

All these projects just address one communication channel; they are not updated anymore and/or need specific hardware. Currently, there is no project that abstracts from the hardware and unifies all different kinds of channels provided by modern smartphone hardware to create a communication layer that is transparent to the user. This fact raises the question, if it is possible to form a mesh network in such environments in general.

III. SIMULATION OF MESH NETWORKS

To answer the question raised in the previous paragraph, the following simulation was implemented.

A. *Simulation set-up*

To evaluate if spanning a mesh network in an environment like a music festival is possible, a simulation of the scenario was conducted using the following steps:

1. Creating a model of the area
2. Generating and distributing nodes
3. Generating edges
4. Analysis

In the first step, a model of the whole area is created. It is divided into sub-areas with varying priorities for the density of people. For example, the area in front of a stage tends to be more crowded than a tent with merchandising. An example for an area with different priorities can be found in Figure 2. The upper left corner will have two times as much people in it and the lower right six times more than the rest of the area.

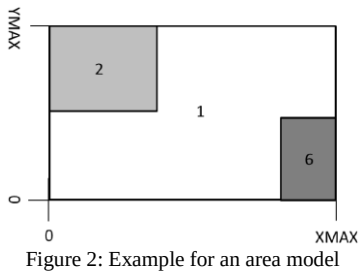


Figure 2: Example for an area model

Second step is to generate nodes which is equivalent to an amount of people. To achieve uniform distribution of the nodes in each subarea, X- and Y- coordinates of a node are represented by random numbers between zero and Xmax/Ymax.

Now, the priorities of each area are used as the probabilities for a Bernouilli experiment [6]. The result of this experiment determines whether or not a randomly generated node is added to the subarea.

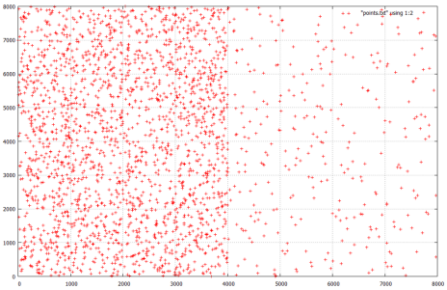


Figure 3: Example plot of distributed nodes

In Fig. 3, the distribution of nodes in an area where the priority on the left is four times as big as in the right half is shown.

After that, the generation of edges between the nodes takes place in the third step. Target of this step is to find out, how many mesh networks could possibly exist within the whole area. To create a mesh network, specific parameters of a communication interface need to be taken into consideration, such as maximum range and maximum number of connections per interface. The range can either be set to a fixed value like the maximum range in the device specification or vary within a certain codomain. The first will create optimistic and the latter pessimistic results. It turns out that the pessimistic results are more realistic, as first tests for Bluetooth pointed out, that with only few people the maximum range of 100m can be reached, but within a huge crowd with lots of devices interfering, it can only be a few meters. Using these values, the approach is as follows:

Starting from a random node, it first will be checked, if this node has reached its maximum connection count. When there are still connections available, the node will be compared to its neighbors. If a neighbor is within the interfaces range and also has connections available, an edge between the two nodes will be created. This will be repeated until there is either no node with free connections left, or all nodes are processed. Output of the third step is an undirected graph.

The final step is to analyze the resulting graph. Using repeated breadth first search until all nodes are marked, it can be determined of how many connected components the graph consists of. This represents the number of possible mesh networks in the modelled area depending on the number of people and specified device parameters.

To receive meaningful results, 22 iterations from 400 to 1500 nodes with increments of 50 were realized. Due to the fact that nodes are placed randomly within the areas, each iteration was executed 10000 times to get good average values. We considered a connection count of seven (active) connections for Bluetooth and an optimistic range of 50m

 ISBN: 978-1-61208-305-6

(half of the specified maximum). The results for a model area of the German music festival "Das Fest" [7] will follow in the next section.

B. Results and Discussion

The probability of connecting all nodes to one single mesh network raises with the number of nodes. Above 1000 nodes, the chance is higher than 90% and gets close to 100% for more than 1300 nodes. Even with only 400 nodes, there are not much more than two separated networks and just around one node without any connection. These numbers lower drastically with more nodes added to the area. A reason, why a full connection of all nodes is not possible is, that the maximum number of connections for each node will be reached at some point.

The values of the optimistic simulation changed drastically if the range is changed to dynamic values between 5m to 100m, depending on the density of people/devices in an area:

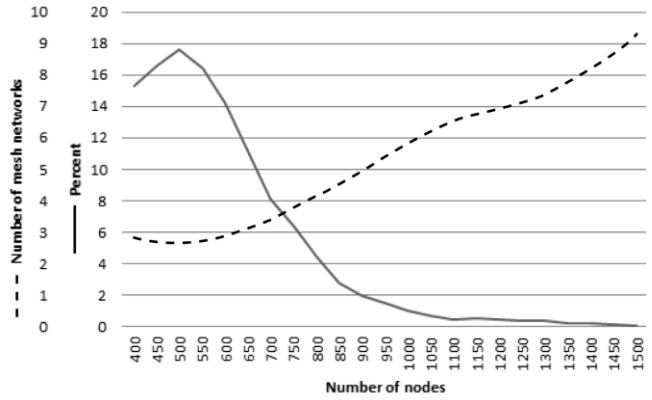


Figure 4: Probability of a single mesh network / Number of networks

As Fig. 4 shows, after a raising (but in comparison still low) probability, it decreases significant for more than 500 nodes and almost reaches 0 for 1200. The number of mesh networks will raise up to 9.5 for a number of 1500 nodes.

The results of both simulations show, that it is possible to create mesh networks within mass events. Even though it is likely to get more than one network, it should be possible to compensate this using only little, well placed infrastructure to connect the different subnets.

This first simulation focused on Bluetooth, but it can be easily adapted to WiFi(-Direct) by changing number of connections and range. Further results are expected within a short timeframe.

IV. A PLATFORM- AND HARDWARE-INDEPENDENT ARCHITECTURE

As shown in related work, there is no real treat to the current issues, but as seen above, the possibility to reach many people on a mass event using mesh networks is given. Therefore, a new architecture for the implementation of mobile mesh networks will be introduced. After pointing out the requirements for such architecture, the different layers will be explained following by an overview of the complete architecture.

A. Claims to the architecture

Goal for the architecture is to abstract hardware and software-platform of mobile devices and enable automatic connection and routing between these. All higher layers should just know, there is a way to communicate, regardless, which specific one it is.

Each component of the architecture should be encapsulated and separated strictly from other components so that the architecture keeps being flexible and customizable without high efforts. The routing e.g. should not correlate with any hardware specific implementations so that a new routing algorithm can be integrated, without touching other code but the router itself.

Finally, the whole architecture and its implementations should be easy to integrate into mobile apps by providing a well-defined interface with just few methods and events.

B. Architecture Layers

The architecture consists of four layers that are independent from each other and only communicate via messages/events and it provides an interface which encapsulates all layers. Thus, each layer can be implemented separately and exchanged by different implementations.

The bottom layer is the datalink layer. It contains all hardware specific code and manages the connections for the different channels. The connector components for each datalink automatically search for available peers and try to connect to them. Once a connection is established, the IO-Stream will be passed to the next layer - the Local Peer Manager.

This layer holds all used datalinks and names for the peers. From here on, the system only deals with names and does not care about which hardware channel is used anymore. Instead it just receives the given streams and forwards them to the Message Broker.

The Message Broker is responsible for parsing incoming byte streams and distinguishing between routing messages and text messages. Routing messages will be deserialized and handed to the routing layer to control the message flow. Text messages will be handed to the router without being touched.

Currently Ad-hoc On-Demand Distance Vector (AODV) [8] routing is used; but, due to the independence of the layers, it could easily be exchanged with Destination-Sequenced Distance Vector (DSDV) [9] or any other routing protocol.

The iWave Communicator surrounds the layers of the whole architecture as a façade. It provides simple functionality to control and reuse the architecture, such as events for new connections and disconnections, listing all available peers as well as methods to send messages or broadcast them to the whole network.

After introducing all layers and components, the following Fig. 5 will provide an overview over the complete architecture and coherences.

 ISBN: 978-1-61208-305-6

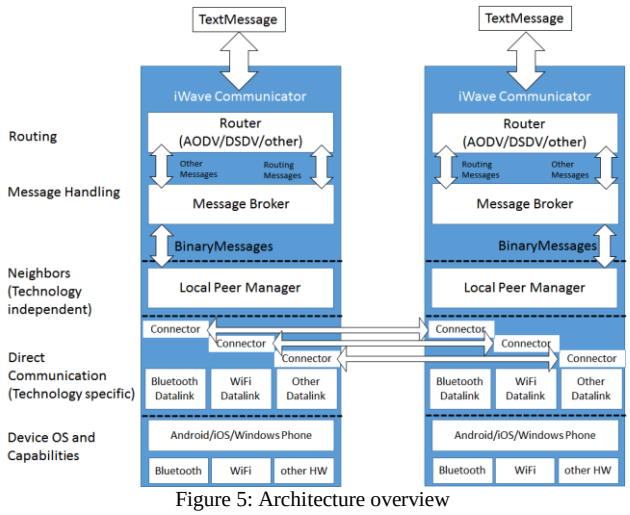


Figure 5: Architecture overview

The given architecture is currently under evaluation and first implementations will be tested soon.

V. CONCLUSIONS

After introducing the need for a hard- and software independent architecture and taking a look at existing approaches it was proven in a simulation, that it is possible to create mesh networks in environments of mass events. Finally, this paper introduced a hard- and software independent architecture to be used on all mobile platforms for this purpose.

First implementations of this architecture on Android and Windows Phone have shown that current mobile platforms are all more or less restricted when it comes to interconnectivity. This makes it hard to achieve the goal of the architecture being totally independent of the platforms.

Android seems to have the fewest restrictions right now. It is able to initiate outgoing as well as accepting incoming connections. Using "InsecureBluetooth" it is even possible to connect two devices without the need of manually pairing the devices, if they are running the same app. For WiFi-Direct there are workarounds to avoid the need for user interaction via hidden API calls and there is also an ongoing discussion of providing an official API call to do so [10].

Windows Phone 7 does not allow to control Bluetooth programmatically at all and WiFi-Direct connections are just allowed outgoing. Windows Phone 8 does not allow any incoming Bluetooth/WiFi connections from non-Windows devices. However, the Phone 8 SDK contains a class called PeerFinder [11] to automate search and connection between two Windows Phone devices at least via Bluetooth. Even though it also contains a property "AllowWifiDirect", this is not supported so far [12].

The possibilities for iOS are still under evaluation and while writing this paper, Apple announced iOS7 [13] which brings a new framework called "Multipeer Connectivity Framework". It is supposed to enable ad-hoc connections between iOS devices. Unfortunately, it seems like this will not work between iOS and other platforms.

VI. OUTLOOK

Due to the mentioned restrictions, further research has to be done to work around these issues and provide connectivity between all different platforms. The vendors should open up their platforms to developers a little more because especially for security related messaging, meshed networks could be ideal. There is clearly the need for a common API and a standardized message format to enable this seamless connectivity across platforms.

After promising results in a lab environment, the first Bluetooth implementation of the proposed architecture was field tested in July 2013 at the music festival "Das Fest" in Karlsruhe, Germany [7] and the collected data is currently being analyzed.

ACKNOWLEDGMENT

Thanks to the Federal Ministry of Education and Research for supporting the project "iWave" (program: "Central Innovation Program SME", ref. no. KF2457707ED2).

REFERENCES

[1] G. K. Still, "Duisburg - 24th July 2010 Love Parade Incident - Expert Report," FIMA Bucks New University, Dec. 2011.

[2] R. Gohs, S. Gunnarsson, and A. Glenstrup, "Beddernet: Application-Level Platform-Agnostic MANETs", in LNCS, Distributed Applications and Interoperable Systems, P. Felber and R. Rouvoy, Eds.: Springer, 2011, pp. 165-178.

[3] Forschungszentrum Informatik, MANET Projekt. Available: http://www.manet-projekt.de Retreived: 10.08.2013.

[4] S. Farahani, ZigBee Wireless Networks and Transceivers. Newton, MA, USA: Newnes, 2008.

[5] K. Panitzek et al., "Can We Use Your Router, Please?: Benefits and Implications of an Emergency Switch for Wireless Routers," Int. Journal of Information Systems for Crisis Response and Management, vol. 4, 2012, pp. 59–70.

[6] J. V. Uspensky, Introduction to mathematical probability. New York etc: McGraw-Hill, 1937.

[7] Das Fest GmbH, Das Fest - Official Website. Available: http://www.dasfest.net. Retreived 08, 2013.

[8] C. Perkins, E. Belding-Royer, and S. Das, Ad hoc On-Demand Distance Vector (AODV) Routing. US: RFC Editor.

[9] C. E. Perkins and P. Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," SIGCOMM Comput. Commun. Rev, vol. 24, no. 4, 1994, pp. 234–244.

[10] Google, Issue 30880: Wi-Fi Direct API for connection acceptance. Available: https://code.google.com/p/android/issues/detail?id=30880. Retreived: 10.08.2013.

[11] Microsoft, PeerFinder Class. Available: http://msdn.microsoft.com/en-us/library/windows/apps/windows.networking.proximity.peerfinder. Retreived: 10.08.2013.

[12] Microsoft, PeerFinder.AllowWiFiDirect. Available: http://msdn.microsoft.com/en-us/library/windows/apps/windows.networking.proximity.peerfinder.allowwifidirect. Retreived:10.08.2013.

[13] Apple Inc, iOS 7 beta for Developers. Available: https://developer.apple.com/ios7/. Retreived: 10.08.2013.

 ISBN: 978-1-61208-305-6

Performance Evaluation on OpenGIS Consortium for Sensor Web Enablement Services

Thiago C. Tavares*†, Regina H. C. Santana*, Marcos J. Santana*, Julio C. Estrella*
*Institute of Mathematics and Computer Science, ICMC/USP
{thiagocp,rcs,mjs,jcezar}@icmc.usp.br
†Federal Institute of Education, Science and Technology of Southern of Minas Gerais, IFSULDEMINAS
{thiago.tavares}@ifsuldeminas.edu.br

***Abstract*—The aim of this paper is to describe a performance evaluation of the interface model of Sensor Web Enablement, especially highlighting the Sensor Observation Service, Sensor Event Service and Sensor Instance Registry. These standards provide a transparent and interoperable way to access data measured by sensors. Studies found in the literature do not treat a performance evaluation on highlighted services in a detailed way. So, the performance evaluation in our study considers several factors that can influence the access time on these services. The results show an important influence of different filter types in the service response times. The result analysis demonstrated that the implementation of application that uses these services should be careful on use of these filters, as, due their definition, the performance of these applications can decrease.**

Keywords*—*Sensor Networks; Service-Oriented Architecture; Web services.

I. INTRODUCTION

A sensor network is composed of sensors that monitor one or a combination of physical data in which the results are sent to an application or final user. It is used in a wide range of monitoring and tracking applications. Furthermore, the breakthrough of their applications has been possible due to the improvement and feasibility of the sensor platforms' cost [1][2]. However, a major challenge in the use of these sensor networks is the feasibility of managing them and providing the necessary information for the use in different applications. On the one hand, there is the infrastructure composed by the sensors and usage strategies of them, as well as the information obtained by them. On the other hand, there are applications or observers who should receive the information and process them. Besides, the sensor networks must also have a communication infrastructure to provide data exchange, between sensors, as well as between network and the observers.

In order to enable the use of sensor networks, it is possible to develop a middleware that provides the tools needed to manage them. Therefore, the literature presents a number of proposals and implementations of middleware used to facilitate the information access provided by these networks regarding the installation, maintenance and execution of applications [3].

One approach that has been proposed in the literature considers the sensor network as a Web Service, i.e., some specifications and languages are used to make an abstraction of the complexity of the sensor system [4]. The abstraction mechanisms provide a standardized interface to access the information following an approach of the Service-Oriented Architectures (SOA). Middlewares that use the SOA concepts have been widely discussed in the literature [5][6]. The OpenGIS Consortium (OGC), a consortium of over 400 companies and academic institutions, has been working on the definition of standards, specifications and programming frameworks in order to use them in the development of sensor networks available as services [7]. In this context, it has been proposed the SWE (Sensor Web Enablement), which is composed of a set of standards, protocols and interfaces that enable the information obtained by the sensor networks to be available through Web Services, following the principles of service-oriented architectures.

Therefore, it is possible to highlight the SOS (Sensor Obervation Service), SES (Sensor Event Service) and SIR (Sensor Instance Registry) services, among the set of interfaces proposed by the SWE. They perform the functions of obtaining observations, alerting and search of sensors, respectively. The SOS is one of the most studied service in the literature, regarding the studies that focus on qualitative and quantitative evaluations on context of SWE service interfaces [8][9][10]. However, there is a gap in relation to a more complete performance evaluation that takes into account other important services, such as the SES and SIR. Thus, this paper presents a performance evaluation that analyzes in detail the main interfaces, defined by SWE, for the access to sensor systems.

This paper is organized as follows: Section II discusses the standards defined in the SWE. Section III presents some works that are related to the one proposed in this paper as well as the gap in the area. Section IV discusses the results of the performance evaluation of SWE services presenting the design of experiments and the evaluation scenario used to perform them. Finally, Section V presents the conclusions and future works that could be developed from the study discussed in this paper.

II. BACKGROUND

As shown in Section I, the OGC is the creator and maintainer of SWE. Since 2003, some work groups have developed and discussed a set of standards that enable the use of sensors exposed through the Web. In this context, sensors are defined as devices that are discovered and accessed through a standardization of protocols and interfaces. They are infrastructures that enable the integration of sensing resources where applications or users can discover, access, modify and register services of alert and sensing, in a standardized way.

 ISBN: 978-1-61208-305-6

Therefore, the WWW provides an infrastructure that enables the sharing of data measured into a sensor system in a well-defined way, abstracting the complexities of the lower layers of the sensing platforms. For example, the standards defined by SWE abstract the details of communication protocols, the hardware architecture and programming languages used in sensor platforms. So, this abstraction facilitates the development of applications. Besides, it allows the developer to concentrate on the logic of its application, not in the details of communication and programming of sensing platforms.

SWE standards are under development and some updates were published in 2012. Bröring et al. [11] presents an overview of these standards and their recent advances and updates. According to the authors, the SWE standards are divided into two informal subgroups: information model and interface model. The former includes data models and encodings used for data representation standards, while the latter comprises different interface specifications of Web services.

Moreover, the information model includes a set of standards that define data models to be used to code the observations of the sensors as well as their metadata. Aiming this, the SWE contains two main specifications: Observation & Measurements (O&M) and the Sensor Model Language (SensorML). The latter specifies a model and a XML codification for describing sensors. In this language, it is mainly defined the location, input and output data, and the phenomena that are observed by sensors. On the other hand, the standard Observation & Measurements defines a framework for the description of the observations made by the sensors. In addition to the standards, other patterns were also defined: the data model (SWE Common) that provides a low-level model for data exchange related to sensors and it is used by several other patterns of SWE. The SWE Common was previously inserted into the SensorML specification, and nowadays, it is available separately as SWE Common 2.0 specification [7].

In turn, the interface model is used to provide a data access mechanism and measurements performed by sensors via a Web service. Several services were defined in the SWE standards, among them it is possible to highlight the SOS, SES and SIR.

A. SOS

The SOS allows obtaining the measured data by the sensors. Besides, it is important to mention that the observations returned by SOS are encoded within the standard O&M. The SOS standard provides an interface to manage and obtain metadata and observations of heterogeneous sensor systems. Thus, this interface defines how the descriptions and observations of sensors are accessed through an interoperable manner. Among the several possible operations by the SOS interface, the following stand out [12]:

- **GetCapabilities**: gets information about the service.
- **DescribeSensor**: gets the description of a sensor or sensor system.
- **GetObservation**: gets a set of observations that may have different filters (time, location, etc.).
- **RegisterSensor**: allows adding new sensors or sensor system in the service.
- **InsertObservation**: allows the addition of new observations for a particular sensor.

B. SES

The SES allows the users registration and/or applications in an alert system. In this case, the user and/or application make the register in the service and receive notifications of it when the criteria for triggering these notifications are met. The SES clients register filters that are used to define the criteria of triggering alerts in a sensor network. Thus, the SES service operates as a Broker of information that carries the mediation between sensor networks and their clients. In general, the notifications made by SES are encoded in the O&M standard. Three levels of filters can be defined in the SES [13]:

- **Level 1**: allows the registration of a filter that sends alerts via an XPath expression.
- **Level 2**: allows the registration of temporal filters, of location and comparison through FES specification (Filter Encoding Specification).
- **Level 3**: allows the determination of filters with multiple patterns. In this case, it is possible to determine a composition of various filters in the emission of alerts.

C. SIR

The SIR provides an interface for managing metadata of sensors. These metadata are encoded through SensorML language. Furthermore, several types of search requests can be submitted to the SIR service. For example, searches can be performed using criteria such as type of service (SOS or SES), types of observed phenomena, location, description, etc. Additionally, it is possible to update sensor information and insert status information of a sensor characteristic as the battery status [14]. The SWE also provides an interface called the SOR for the management of the semantics of the phenomena observed by the sensors. However, this service is not addressed in the study presented in this paper. Section III presents some related works and the gaps identified in these studies.

III. Related Work

This section aims to present some works related to qualitative and quantitative evaluations in the context of the SWE standards. The work presented by McFerren et al. [8] discusses implementations of the Observation Service Sensor highlighting features such as easy installation, documentation quality, and completeness of implementation in relation to the standard definitions. The authors consider four types of implementations: 52°North Initiative, PySOS, MapServer and Deegree SOS and they do not consider any quantitative analysis such as a performance evaluation of the implementations concerning the functionalities provided by them.

Moreover, Poorazizi et al. [10] presents a complementary study of the work found in [8]. The performance of several implementations of SOS services. The authors present a

 ISBN: 978-1-61208-305-6

review of SOS considering different filters of data acquisition such as number of sensors, location, and time. The study has considered three of the four implementations discussed [8] (52°North, Deegree, and MapServer). Furthermore, the performance analysis took into account two characteristics: response time and size of documents returned by the service.

In turn, Tamayo et al. [15] presents a performance evaluation of SWE standards in a mobile computing environment. In this study, the authors evaluated the performance of different Smartphone in the document processing with sensor observations obtained through SOS. Besides the processing, the authors also considered the size of these documents and their transmission through different types of networks such as Wi-Fi and 3G, as well as different XML processing APIs for the Android platform.

Finally, Tamayo et al. [9] presents an empirical study of current instances of SOS providers. The authors conducted an investigative work raising tens of SOS services available on the Web. These services have undergone several tests to check, for example, which parts of the specification are more frequent in SOS service implementations. Besides, the authors also found that many of the implemented providers have validation problems with the documents of observations returned by these servers, i.e., many of the documents returned by these servers could not be validated with the XML Schema that defines them.

As shown in this section, several studies in the literature analyze the SOS service, although many other services of SWE interfaces model are not considered. For example, SES is an important service within the interfaces model and it has not been treated by the literature in studies of performance evaluation. Alert services are important tools for developing applications of critical systems, which the delays in the delivery of alerts can hinder the effectiveness of these applications. Additionally, the registry service (SIR) is not considered in others SWE performance evaluation studies. The SIR is an important discovery service of sensor systems, although it is not a pattern of SWE yet. Currently, the SIR is treated as a "discussion paper". However, it is already possible to find available implementations of this service as the one available on the website of 52° North [16]. Thus, Section IV aims to present and discuss the methodology and results of a performance evaluation of the SWE interfaces model, especially regarding the services SOS, SES and SIR.

IV. Performance Evaluation

This section aims to present a performance evaluation of SOS, SES and SIR services that compose the model of the SWE interfaces. Therefore, the purpose of this evaluation is to verify distinctions in performance using different types of filters in requests submitted to these services. Additionally, the evaluation proposed in this section considers a full factorial experiment design with three factors and two levels: **Amount Of Clients**, **Submitting Rate** and **Filter Types** (2^3, 8 Experiments). This design is applied to each of the evaluated services and it is defined in Table I.

The Amount of Clients and Submitting Rate factors possess the same levels for all services evaluated. The variation in the number of clients is performed by creating multiple threads that mimic the behavior of multiple clients accessing the services. In turn, the Submitting Rate factor simulates the submission of requests rate following an exponential function with averages of 120 and 240 requests per minute. Besides, it is important to know that each client (thread) submits 10 requests to the service using the exponential function highlighted.

TABLE I. Experiment Design

	Amount Of Clients	Submitting Rate	Filter Types
SOS	50/100	120/240	1Obs/288Obs
SES	50/100	120/240	Level1/Level2
SIR	50/100	120/240	Phenomenon/ID

The Filter Types factor has different levels, respecting the specificity of each service. In the SOS service case, are tested two variations of the GetObservation requests. The SOS services configured on the machines contain a database with the observations of sensors that measure the level of water concentration. The insertion of the observations in the database mimics the behavior of a sensor network by sending an observation every 5 minutes to the SOS service during a month. This behavior generates a total of 8640 observations registered in the server of the service provider. Therefore, in the context of the SOS experiments, the variations in the request messages are in relation to the periods of time to obtain the observations. The first experiment of SOS service concerns a period of time, which only one observation is returned, while the second type takes into account a period that the observations of a day are returned, totaling 288 observations.

In the case of the SES service, Level 1 and Level 2 that define the criteria for triggering alerts are used. As mentioned in Section II-C, the Level 1 considers a XPath expression that checks the value of the element *om:procedure*, while the Level 2 takes into account criteria such as sensor location, value observation, etc. In the case of the experiments performed in this performance evaluation, it is considered a criterion for location shooting, i.e., there will be an alert triggering when the SES receives sensors data that are located in a certain area. Finally, the experiments performed by the SIR consider two types of search criteria: the name of an observed phenomenon and the ID of the sensor in the service registry. The configuration of the SIR for this evaluation has 12 registered sensor systems that offer the same sensing information. Thus, experiments using a filter for the name of the phenomenon return 12 sensors descriptions (SensorML). However, the use of the ID in the search filter returns only one description. Section IV-A presents the infrastructure and the scenario implemented to perform the experiments.

A. Evaluation Scenario

The evaluation scenario uses an infrastructure composed of two virtualized machines (KVM) on different physical nodes. The physical nodes used for virtualization of these two machines have the following characteristics:

- Processor: Intel(R) Core(TM)2 Quad CPU Q9400 of 2.66GHz.
- Memory: 8 GB RAM DDR 3.
- Size disk: 500 GB. 7200 RPM.

 ISBN: 978-1-61208-305-6

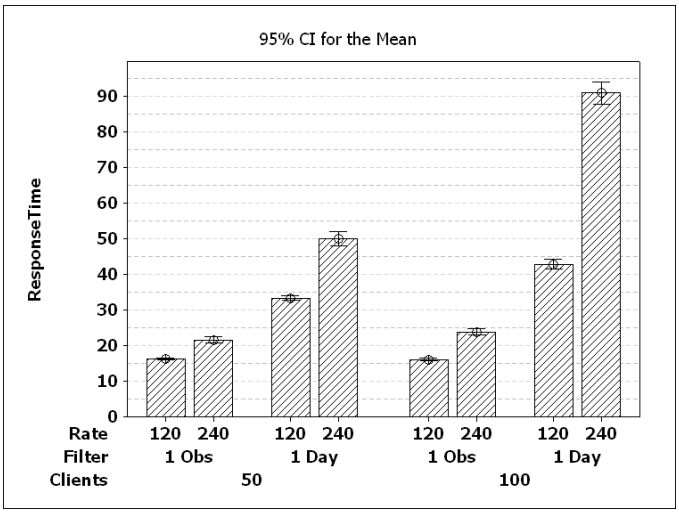


Figure 1. SOS: Response Times

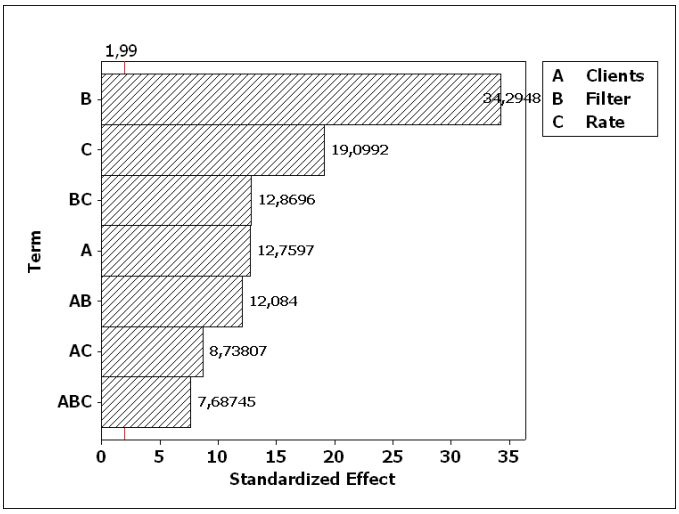


Figure 2. SOS: Factor Influence

In turn, the two virtual machines instantiated for the experiments have different settings, following the characteristics defined in Table II.

TABLE II. VIRTUAL MACHINE SETTINGS

Machine	Processors	Memory	Disk Size	Operating System
Server	4	4GB	15 GB	Ubuntu 12.04 (64-bits)
Client	2	2GB	15 GB	Ubuntu 12.04 (64-bits)

Regarding software, it was used the implementations provided by 52° North Initiative. It was used the versions as follows [16]:

- SOS: 3.5.0 version;
- SES: 1.0.0 version;
- SIR: 0.4 version;

B. Results

The results of the design of experiments presented in this section are shown in two types of charts:

- **Charts of the response times**: in these charts are presented the variations of the average response times in relation to variation in the levels of the factors. The confidence intervals calculated use a 0.05 alpha (95% of confidence). Furthermore, the averages are obtained by performing 30 replicates for each experiment.
- **Pareto Charts**: these charts show the influences of each of the factors in the tests. They use a vertical line that indicates the point where the factors start to have an influence in the experiments. In other words, the factors that lie above that line influence the response time. Additionally, the calculation of the influence percentage of each factor can be achieved through of calculating of each value of the factors in the Pareto chart divided by the sum of all of them.

As mentioned in Section III, several works performed studies of SOS services performance. However, the experiments conducted about this service in the study presented in this paper differ from those found in the literature. The performance evaluations on the SOS presented here use different evaluation factors. Besides, the analysis considers the behavior of the SOS service in relation to the variation of the number of clients accessing the service and the request rate submitted by each of them, in addition to the filters that determine different amounts of returned values. Thus, the chart in Figure 1 shows that the largest increase in response times occurs on changing the filter that returns only one observation for a filter that returns 288 observations (1 day of observation). In other words, significant increases in response times, considering the increase of clients, occur to the filter of 1 day. Response times are close in relation to the increase of clients for experiments with requests that return only 1 observation. The Pareto chart in Figure 2 shows that all factors influence the response time in the experiments, including the interactions between factors. In summary, the Filter factor has 31.9% of influence followed by Submitting Rate with 17.8% and Amount of Clients with 11.9%. Although the type of filter used has a greater impact on the response times, it is important to consider the number of customers and the rate of submission of requests, mostly for filters that return many observations.

The results obtained for the SES are shown in Figures 3 and 4. In the Figure 3, it is possible to observe that the large difference in response times occurs when the amount of clients are differents. Additionally, levels of filters also influence on the response times, especially for experiments with the average of 240 requests per minute and the experiments with 100 clients. In such cases, the experiments that consider the Level 2 have response times considerably higher than those obtained by the Level 1. The Pareto chart shown in Figure 4 shows that the number of customers is the most prominent factor in the experiments, followed by the factors of filters level and rate of requests. Therefore, the Amount of clients factor has an influence of 24.5% approximately, whereas the filter and rate factors have an influence of 21.9% and 18%, respectively. You can also verify that the interaction between these factors also represents significant influences. One of the

 ISBN: 978-1-61208-305-6

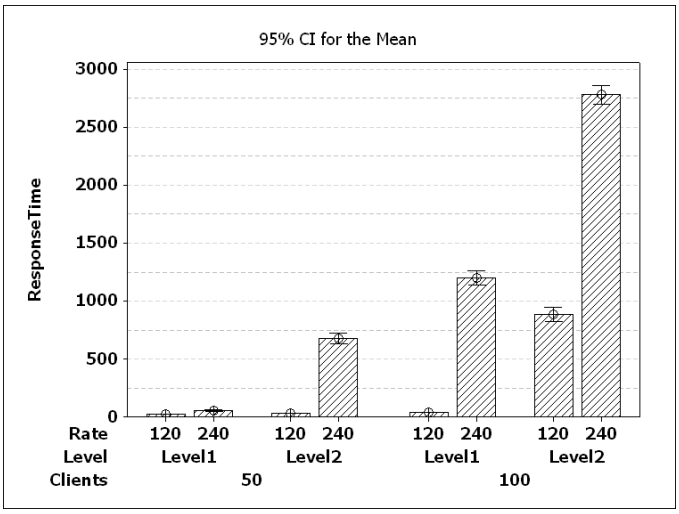


Figure 3. SES: Response Times

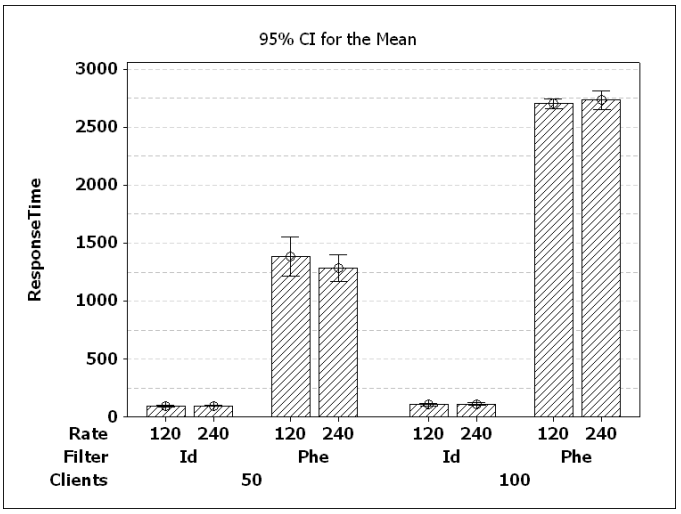


Figure 5. SIR: Response Times

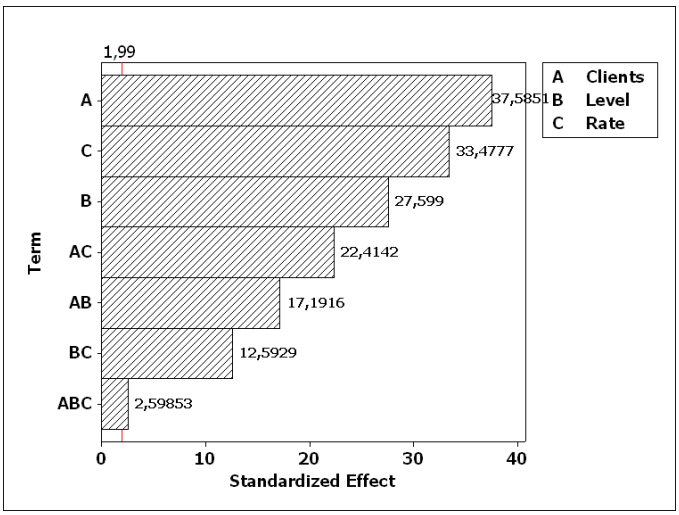


Figure 4. SES: Factor Influence

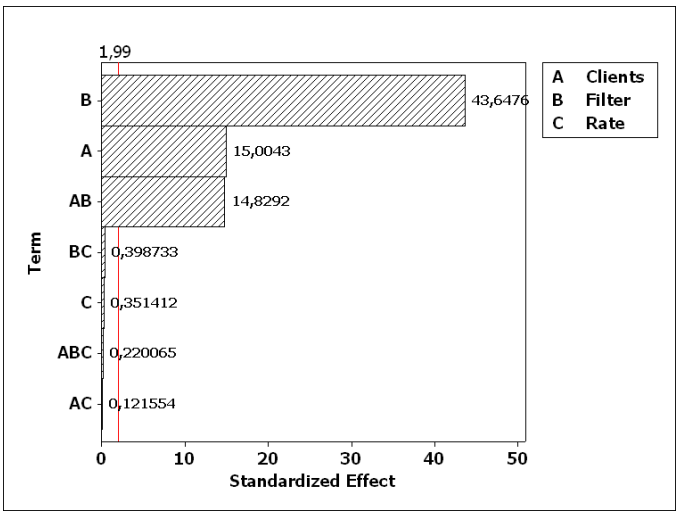


Figure 6. SIR: Factor Influence

main findings obtained in the execution of the experiments is related to the influence of the filters levels used in SES service. Applications that use SES can employ the results obtained in the experiments of this service to mark the usage of filter types in a more rigorous way. For example, certain applications that receive data from sensors networks and which react to alert messages may opt to computationally lighter filters as in the case of Level 1, when possible. Thus, as shown in the experiments, the proper definition of the filters can improve performance in the process of alerting.

Finally, the experiments related to the SIR are shown in Figures 5 and 6. The chart in Figure 5 shows that response times have significant differences in the Filter factor. Besides, the search for information of sensors using its ID in the service is much more efficient, since there is only one description of the sensor. However, it is impossible to know the ID of the sensor without performing a more generic search, such as the name of the observed phenomenon. Thus, if the application needs to check frequently possible updates in the sensor description, it firstly uses a search for the observed phenomenon and the subsequent searches by the ID obtained in the first interaction. Another mechanism that may be used to optimize the search of sensor systems in the SIR is the insertion of a broker that makes a cache of the search messages sent to the SIR. In this case, the broker can relate the search messages with the sensor Ids returned by SIR. Thus, the Broker can use the identifiers through search messages stored in the cache. For example, a client does a search for sensors that measure the wind speed and submits this search to the Broker. Then, the Broker receives this search message and forwards it to the SIR. The SIR response is stored in a tuple with the search message and sensor ID (*find_msg,sensor_id*) in the cache Broker. Therefore, when other clients submit the same search message, the broker replaces this message by a search message through the sensor ID, reducing the access time to the service registry.

Furthermore, searches performed by the ID of the sensor have no significant changes in time with the increase of clients' number and the rate of submitting requests. In such cases, it is possible to observe that the averages are statistically equal.

 ISBN: 978-1-61208-305-6

This behavior is reflected in the Pareto chart, demonstrated in Figure 6. The chart shows an influence of 58.5% in the filter factor.

The results presented in this paper demonstrate a performance differences on distinct types of filters in the considered services. So, the appropriate choice of these filters can benefit the performance of applications that use the SWE standards. For example, the developers of SWE applications have a better option in filters choosing that return less data. In high workload situations, the response time on changing a filter that returns only one observation for a filter that returns one day of observation can increase almost three times. In turn, the levels of filters on SES services also influence the performance of applications that use this service. Applications most rigorous regarding response time should choose level 1 filter that have better results and do some value comparison on own logic. Finally, searches on SIR hold improved performance using filter by ID. However, it is impossible to know an ID without using another filter type. So, it is indicated the use a phenomena filter, for example, in first search and a search for ID for the other searches. This type of interaction is indicated to application that send several searches for same ID to verify changes on descriptions of the sensor systems.

The results also show important influences in factors as amount of clients and submitting rate. They impact the response time in several tests. A solution to improve the performance of applications respecting these factors should be a cloud infrastructure. In this case, it is interesting to have an infrastructure where is possible to increase the computational capacity that offers the service. The OGC mentions the use of a cloud infrastructure in a white paper published in its official site [17]. Section V presents the performance evaluation conclusions and it discusses future works that can be developed from this study.

V. Conclusion and Future Work

This paper presented a performance evaluation of the interface models of SWE, especially highlighting the SOS, SES and SIR services. This evaluation considered the amount of clients, type of filters and submission rate as influencing factors in response times when accessing the services highlighted. Therefore, the results demonstrated an important influence of the filters type in the service response times. The influence of different filters in the requests was 24.5%, 31.9% and 58.5% for the SES, SOS and SIR services, respectively. The analyzes showed that the implementation of applications that use these services should carefully use the filters of these services, since the definition of them can significantly impact the performance of these applications.

Future studies should be developed to consider other services of SWE as SPS, and also improve the performance evaluation by increasing the variation of these filters. In the case of SIR Service, a Broker that manages the search messages to optimize the performance in accessing this service can be developed. Moreover, it is possible to develop mechanisms in relation to the provision of quality of service in the access of SWE interfaces model services, once the patterns specified do not consider this type of problem.

Acknowledgements

The authors would like to thank the financial support of FAPESP (São Paulo Research Foundation), FAPEMIG (Minas Gerais Research Foundation), CAPES (Coordination for the Improvement of Higher Education Personnel) and IFSULDEMINAS (Federal Institute of Education, Science and Technology of Southern of Minas Gerais).

References

[1] I. Akyildiz and M. C. Vuran, *Wireless Sensor Networks*. New York, NY, USA: John Wiley & Sons, Inc., 2010.

[2] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, 2008.

[3] M. Wang, J. Cao, J. Li, and S. K. Das, "Middleware for wireless sensor networks: A survey," *J. Comput. Sci. Technol.*, vol. 23, no. 3, pp. 305–326, 2008.

[4] T. Laukkarinen, J. Suhonen, and M. Hännikäinen, "A survey of wireless sensor network abstraction for application development." *IJDSN*, vol. 2012, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/ijdsn/ijdsn2012.html#LaukkarinenSH12

[5] F. C. Neto and C. M. F. A. Ribeiro, "Dynamic change of services in wireless sensor network middleware based on semantic technologies," *Autonomic and Autonomous Systems, International Conference on*, vol. 0, pp. 58–63, 2010.

[6] N. Mohamed and J. Al-Jaroodi, "Service-oriented middleware approaches for wireless sensor networks," in *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, 2011, pp. 1–9.

[7] OGC, "Ogc standards and specifications," December 2013, available in: http://www.opengeospatial.org/standards. Last Access: 06/05/2013.

[8] G. McFerren, D. Hohls, G. Fleming, and T. Sutton, "Evaluating sensor observation service implementations," in *Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009*, vol. 5, 2009, pp. V–363–V–366.

[9] A. Tamayo, P. Viciano, C. Granell, and J. Huerta, "Empirical study of sensor observation services server instances," in *Advancing Geoinformation Science for a Changing World*, ser. Lecture Notes in Geoinformation and Cartography, S. Geertman, W. Reinhardt, and F. Toppen, Eds. Springer Berlin Heidelberg, 2011, pp. 185–209. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-19789-5_10

[10] M. E. Poorazizi, S. H. L. Liang, and A. J. S. Hunter, "Testing of sensor observation services: a performance evaluation," in *Proceedings of the First ACM SIGSPATIAL Workshop on Sensor Web Enablement*, ser. SWE '12. New York, NY, USA: ACM, 2012, pp. 32–38. [Online]. Available: http://doi.acm.org/10.1145/2451716.2451721

[11] A. Bröring, J. Echterhoff, S. Jirka, I. Simonis, T. Everding, C. Stasch, S. Liang, and R. Lemmens, "New generation sensor web enablement," *Sensors*, vol. 11, no. 3, pp. 2652–2699, 2011. [Online]. Available: http://www.mdpi.com/1424-8220/11/3/2652

[12] SOS, "Ogc - sensor observation service interface standard," 2012, available in: http://www.opengeospatial.org/standards/sos. Last Access: 3/06/2013.

[13] SES, "Opengis - sensor event service interface specification," 2008, available in: http://portal.opengeospatial.org/files/?artifact_id=29576. Last Access: 12/06/2013.

[14] SIR, "Sensor instance registry discussion paper," 2010, available in: http://portal.opengeospatial.org/files/?artifact_id=40609. Last Access: 25/05/2013.

[15] A. Tamayo, C. Granell, and J. Huerta, "Using swe standards for ubiquitous environmental sensing: A performance analysis," *Sensors*, vol. 12, no. 9, pp. 12 026–12 051, 2012. [Online]. Available: http://www.mdpi.com/1424-8220/12/9/12026

[16] 52North, "Softwares," 2013, available in: http://52north.org/downloads/sensor-web. Last Access: 20/08/2013.

[17] SOS, "Ogc standards and cloud computing," 2011, available in: https://portal.opengeospatial.org/files/?artifact_id=43743. Last Access: 20/08/2013.

 ISBN: 978-1-61208-305-6

Commercialized Practical Network Service Applications from the Integration of Network Distribution and High-Speed Cipher Technologies in Cloud Environments

Kazuo Ichihara, Naoko Nojima
Net&Logic Inc.,
Meguro-ward,
Tokyo, Japan
e-mail: ichihara@nogic.net, nojima@nogic.net

Yoichiro Ueno, Shuichi Suzuki, Noriharu Miyaho
Department of Information Environment,
Tokyo Denki University, Inzai-shi,
Chiba, Japan
e-mail: ueno416@mail.dendai.ac.jp,
suzuky@mail.dendai.ac.jp, miyaho@mail.dendai.ac.jp

Abstract— **This paper presents the evaluation results of the commercialization of High Security Disaster Recovery Technology (HS-DRT) that uses network distribution and high-speed strong cipher technologies to realize efficient and secure network services. We have commercialized a disaster recovery system and evaluated the performance of the distributed engines using the hash functions, versatile spatial scrambling functions, etc., in cloud computing environments. The average processing time has been estimated in terms of the method of implementation of the engine. As for practical network applications, an automatic back-up system using an FTP server has been introduced. We have developed on-premise systems which achieve high security through the use of HS-DRT. Finally, we also propose future technologies for preventing an insider attack.**

Keywords-disaster recovery; backup; distributed processing; cloud; strong cipher.

I. INTRODUCTION

Innovative network technology, which can guarantee, as far as possible, the security of users' or institutes' massive files of important data from any risks such as an unexpected natural disaster, a cyber attack, etc., are becoming increasingly indispensable day by day. As a means of satisfying this need, cloud computing technology is expected to provide an effective and economical backup system by making use of a very large number of data stores and processing resources which are not fully utilized. It is expected that this file data backup mechanism will be utilized by government and municipal offices, hospitals, insurance companies, etc., to guard against the occurrence of unexpected disasters such as earthquakes, large fires and storms and Tsunamis. To achieve secure back up, there is an indispensable need for prompt restoration, which may make versatile use of cellular phones, smart phones, digital signage equipment and PCs, in addition to cloud resources dispersed in multiple geographical locations. In addition to these factors, many companies and individuals involved in industry and commerce are interested in making use of public or private cloud computing facilities, provided by carriers or computer vendors as a means of achieving security and low maintenance and operation costs. In this paper we present the results of an evaluation of an innovative file backup network service, which makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed strong cipher technology to realize efficient, safe data backup at an affordable maintenance and operation cost [1-6].

When a block cipher is used, the required processor and memory costs increase in an exponential manner with increasing data volume. However, with a stream cipher, the input data is simply operated on bit-by-bit, using a simple arithmetic operation, and high-speed processing becomes feasible. This is the fundamental difference between the two cipher technologies. It is possible to combine the use of technologies, specifically, the spatial scrambling of all data files, the random fragmentation of the data files, and the corresponding encryption and replication of each file fragment using a stream cipher. Figure 1 shows the concept of the proposed network service compared with a conventional back up system using the leased lines.

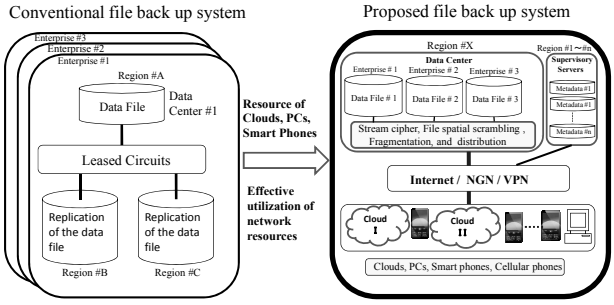


Figure 1. Concept of the proposed network service

In the data center, it is appropriate to introduce a secret sharing scheme for sending "encryption metadata" to the supervisory servers deployed in the several different locations for deciphering the original file data. To enhance security it is better to send the metadata using a Virtual Private Network(VPN). This mechanism make it quite difficult to find out any series in the "encryption metadata" itself. From a disaster recovery point of view, a secret sharing scheme with some appropriate "thresholds" should be introduced in the proposed system. If the system uses a (3,5)-threshold scheme, the system uses five supervisory servers, and can tolerate the simultaneous failure of two servers. On the other hand, from a cyber terrorism point of

 ISBN: 978-1-61208-305-6

view, if the system uses a (3, 5)-threshold scheme, a cracker has to intrude at least three "encryption metadata" servers and one alive/valid information server at the same time [1].

We have developed a high security disaster recovery technology (HS-DRT) for realizing file backup network services. To realize the proposed DRT mechanism, the following three principal network components are required: (1) a secure data center, (2) several secure supervisory servers, and (3) a number of client nodes such as smart phones, cellular phones, digital signage equipment or cloud storage which are available for use. The corresponding history data, including the encryption key code sequences, which we call "encryption metadata", are used to recover the original data. This mechanism is equivalent to realizing a strong cipher code, comparable to any conventionally developed cipher code, by appropriately assigning affordable network resources.

To realize a safe and highly secure system, it is necessary to assure the users that their important file data cannot be stolen from the service provider. When the important file data is composed of a number of fragments distributed to the several providers' clouds, we need to ensure that even a single fragment cannot be deciphered by any of the providers. From this point of view, we considered the special case of preventing an insider attack. The proposed technology can also increase both the cipher code strength and the operation of decryption and reassembly of original file data.

In this paper, we briefly describe related work in Section II, and then, the basic configuration of the HS-DRT engine in Section III, and a performance evaluation of the proposed network services using the HS-DRT engine is presented in Section IV. Practical commercialized systems are discussed in detail in Section V. In Section VI, we describe the technique behind the proposed method of preventing insider attacks. Finally, we describe the conclusions and the future issues in Section VII.

II. RELATED WORK

In the field of Disaster Recovery Systems, there have been many research publications and many commercial products. Most disaster recovery systems include data replication functions using stand-by servers in remote locations. In contrast, the proposed disaster recovery system using HS-DRT uses a secure distributed data backup scheme. By making use of the HS-DRT mechanism to achieve a reliable backup scheme, we have been able to provide a system product at a reasonable price to both individuals and companies. For example, we can provide only the backup application by using HS-DRT with multiple cloud services as backup storage in accordance with appropriate customer contracts. So, it is rather difficult to compare our proposed system quantitatively with an ordinary disaster recovery system from the viewpoint of the price.

In the field of secure data backup systems, other related studies have included the concept of a distributed file backup system [7][8]. However, in these studies, neither a precise performance evaluation nor a practical network service system is clearly described.

In the field of intrusion tolerance, a file server should introduce such functions as encryption, fragmentation, replication, and scattering [9]. The core technologies of HS-DRT resemble those of a persistent file server, except for the spatial scrambling and random dispatching technology. With these two technologies, deciphering by a third party, by comparing and combining the encrypted fragments, becomes almost impossible. In addition, HS-DRT is applicable to other fields, such as secure video streaming, etc.

In the field of Distributed Anonymous Storage Services, the replication and scattering techniques were introduced in the Eternity Service [15]. However, the main objectives of these services are longevity and anonymity.

In contrast, these objectives (longevity and anonymity) are not taken into account as primary requirements in the HS-DRT. HS-DRT is effectively utilized for the purpose of fast, safe and secure file back up until the next backup event. Only the authorized users are able to initiate the process of recovery of their file data contents.

III. BASIC CONFIGURATION OF THE HS-DRT ENGINE

The HS-DRT file backup mechanism has three principal components, as shown in Figure 2. The main functions of the proposed network components, which are Data Center, Supervisory Server and various client nodes, can be specified as follows. The client nodes (at the bottom of Figure 2 are PCs, Smart Phones, Network Attached Storage (NAS) devices, Digital Signage and Storage Services in the Cloud. They are connected to a Supervisory Server in addition to the Data Center via a secure network.

The Supervisory Server (on the right in Figure 2) acquires the history data, which includes the encryption key code sequence (metadata) from the Data Center (on the left in Figure 2) via a network. The basic procedure in the proposed network system is as follows.

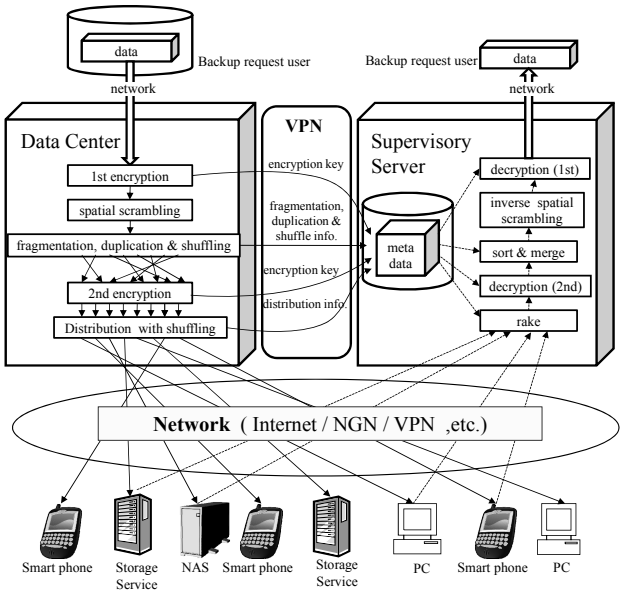


Figure 2. Principle of HS-DRT file backup mechanism

A. Backup sequence

When the Data Center receives the data to be backed up, it encrypts it, scrambles it, and divides it into fragments, and thereafter replicates the data to the extent necessary to satisfy the required recovery rate according to the pre-determined service level. The Data Center encrypts the fragments again in the second stage and distributes them to the client nodes in a random order. At the same time, the Data Center sends the metadata to be used for deciphering the series of fragments to the supervisory servers. The metadata comprises encryption keys (for both the first and second stages), and several items of information related to fragmentation, replication, and distribution.

B. Recovery sequence

When a disaster occurs, the Supervisory Server initiates the recovery sequence. The Supervisory Server collects the encrypted fragments from the various appropriate clients in a manner similar to a rake reception procedure. When the Supervisory Server has collected a sufficient number of encrypted fragments, these are decrypted, merged, and descrambled in the reverse order of that performed during the second stage of encryption, and the decryption is then complete. Through these processes, the Supervisory Server can recover the original data that has been backed-up.

Let us consider the probability of successful recovery, which can be estimated from the following equation. Here, P is the failure rate of each client node, n is the degree of duplication of each fragment, and m is the number of fragments [1].

$$\text{Probability of recovery} = (1 - P^n)^m \cong 1 - mP^n$$

$$\text{Probability of recovery failure} \cong mP^n$$

For example, when each file fragment's failure rate P is assumed to be 0.2, and the original file is divided into 30 fragments, and 30 replications are made of each fragment, the probability of recovery failure becomes less than 10^{-19}. The above case applies to the use of smartphones, cellular phones, or PCs.

The failure rate of such devices can be estimated to be 0.2 by considering their connectivity and reliability, erring on the safe side. Here, the size of users' important data is classified into three types, called Type1, Type2, and Type3. The data size of Type1 is at most around several hundreds of megabytes in size. The data for Type2 is at most around several tens of gigabytes, while that for Type3 is up to several terabytes. Considering a smart phone's memory capacity to be 32 Gbytes, and the number of terminals to be several tens of millions, these are used for Type1 and Type2 data, with the assurance that less than 1% of the vacant memory resource in the terminals offered would be used to support the backup service. This percentage can usually be measured by the user's self-check monitoring and the monitored result can be transmitted to the remote supervisory center. The value of 1% is an example of the conditions to be temporarily assigned to encourage people to participate. Several cloud storage resources can be effectively utilized for Type3 data. When cloud storage is used, then P can be less than 10^{-11} and a reliability much higher than that available commercially can be easily obtained.

The security level of the HS-DRT does not only depend on the cryptographic technology but also on the method of specifying the three combined factors, that is, spatial scrambling, fragmentation/replication, and the shuffling algorithm. Because of these three factors, nobody is able to decrypt the data without collecting all relevant fragments and sorting the fragments into the correct order. Even if some fragments are intercepted, nobody is able to decrypt parts of the original data from such fragments.

The spatial scrambling procedure can be realized by executing a simple algorithm using a C-style description [1]. This computation process should be repeated several times. To de-scramble, it is only necessary to perform the same operations in the reverse order. Introducing the above mentioned spatial scrambling technology makes it almost impossible for a third party to decipher the data by comparing and combining the encrypted fragments, since the spatial scrambling scatters the relevant fragments widely and uniformly amongst the storage devices.

One of the innovative ideas of HS-DRT is that the combination of fragmentation and distribution can be achieved in an appropriately shuffled order. Even if a cracker captured all the raw packets passing between the data center and the client nodes, it would be extremely difficult to assemble all the packets in the correct order, because it would be necessary to try about N! (N: number of fragments) possibilities, where N is sufficiently large. In fact since the bit patterns of any two encrypted fragments are completely different from each other owing to the different encryption keys, it is impossible to associate one encrypted fragment with another. Crackers would require innumerable attempts to decipher the data. In addition, HS-DRT mainly uses a shuffling method that uses pseudo-random number generators for the distribution to the client nodes. When we distribute the fragments of the encrypted data to widely dispersed client nodes, we can send them in a shuffled order, since we predetermine the destination client nodes from the shuffled table in advance. When we use a shuffle table which makes use of the "Fisher-Yates shuffle" algorithm with 3 rounds, leading to a uniform distribution, the table itself is hard for a third party to guess [10].

Practical systems to realize a hybrid HS-DRT engine can be realized effectively by making use of a cloud computing system at the same time. The system essentially consists of the following four parts: thin clients, a web applications server (Web-Apps Server), an HS-DRT engine, and Storage Clouds. Thin Clients are terminals which can use web applications in a SaaS (Software as a Service) environment. Thin Clients can make use of the application services which are provided by the web applications server. The HS-DRT engine is considered to be a component of the hybrid cloud computing system, which can also strengthen the cloud computers' security level at the same time. Usually, users can also make use of one of multiple HS-DRT engines available in the cloud environments through a contract with the providers. The data center and the Supervisory Server

 ISBN: 978-1-61208-305-6

can be integrated in the HS-DRT engine. The HS-DRT engine can effectively utilize the storage clouds which have a function related to a web application and execute the encryption, spatial scrambling, and fragmentation of the corresponding data files. It is very important to note that the processing efficiency of the HS-DRT engine can easily be improved by increasing the amount of the web cache memory.

We need to consider the scalability of the HS-DRT engine, since it may become a bottleneck in a very large system, owing to the number of clients and the amount of storage. In such cases, the HS-DRT engine may use a key-value database. As the HS-DRT engine can easily work with other HS-DRT engines, the system can be extended.

IV. PERFORMANCE EVALUATION OF THE HS-DRT

To study the implementation of the spatial scrambling functions in the HS-DRT engine, we evaluated a simple non-optimized method, and an optimized method for use with multi-core processors; for each method we used four sizes of data, and the whole simulation was written in the C language. The original data is divided into 64-kbyte data blocks and each 64-kbyte data block is further divided into 64 data blocks. The resulting 1024-byte data blocks are shuffled using the Fisher-Yates algorithm. Since the scrambling effect is limited within the range of 1024 bytes, the 64 1024-byte data blocks are further shuffled three times using the Fisher-Yates algorithm in order to achieve the appropriate degree of randomness.

The above mentioned data processing has been implemented using a single thread as a basic application program. To optimize for multi-core processing, the Fisher-Yates shuffling is applied to each thread in the environment of the ubuntu12.04LTS OS.

We adopted three types of CPU, which were 1-core (AMD Athlon 1640B), 2-core (Intel Celeron G530), and 4-core (Intel Corei7), from the viewpoint of economy. We evaluated the processing time by using data sizes of 64kbytes, 640kbytes, 6.4Mbytes and 64MBytes. The measured results are shown in Figure 3. In the graphs, the processing time is shown in terms of the equivalent bit-rate. In case of the 1-core processor, the efficiency of the single thread processing for each 64-kbyte data block (referred to as the 64k-single method) is approximately the same as the case of multi-thread processing for each 64-kbyte block (referred to as the 64k-multi method). It shows that there is no benefit in multi-processing in the case of 1-core processing. In contrast, in the case of the 2-core and 4-core processing, when handling 6.4 Mbytes of data, the speed of the 64k-multi method is five times faster than the 64k-single method for 2-core processing, and nine times faster than the 64k-single method for 4-core processing. Consequently, we derived the following two characteristics of the HS-DRT engine.

1) The 64k-multi method can be used effectively with multi-core processing in handling the spatial scrambling and shuffling procedure under the Linux OS.

2) Since an increase in scrambled data size does not result in much increase in the time required by the Fisher-Yates shuffling, it is recommended to increase the size of the data block for spatial scrambling by utilizing the 64k-multi method.

64K-single method [msec]

	64K	640K	6.4M	64M
1Core	16	96	902	9004
2Core	8	48	440	4320
4Core	4	42	372	3604

64K-multi method [msec]

	64K	640K	6.4M	64M
1Core	16	94	924	9012
2Core	1	10	88	853
4Core	1	10	42	412

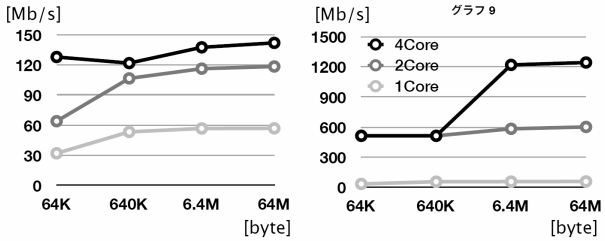


Figure 3. Evaluation of HS-DRT

V. PRACTICAL COMMERCIALIZED SYSTEM

We have considered the following business model.

1) The proposed system can share the network resources such as clouds for different users with different security and availability requirements. This can lead to an effective cost reduction compared with conventional systems using leased lines and data centers.

2) The proposed system can utilize the enormous number of PCs, smart phones, and digital signage systems as network resources to provide the backup services in accordance with individual contracts with the owner users. This also can lead to an effective cost reduction compared with the conventional network system provider resources.

3) By effectively making use of the above mentioned network resources, it is possible to provide backup services for only a little bit more than the cloud usage prices.

4) For users who cooperate in offering available unused network resources (memory), we can offer them a communication tariff reduction or alternatively free backup services.

A. Simple FTP server automatic backup system without using Clouds

In 2010, we commercialized the personal HS-DRT backup system that is called "@Cloud-DRT backup". In this system, the customer only installs client software on their PC, and drags-and-drops the required files to the specific folder. This process is suited to personal use, but it is not fit for a large scale file service, such as an FTP server.

We have therefore made an autonomous backup system for FTP servers by adapting the way "@Cloud-DRT backup" is used. The components of our autonomous backup system are shown in Figure 4. We added a backup server and only installed the watch program on the target ftp server, without modification of the ftp daemon. When a user uploads a file to the FTP server, the watch program needs to look for the

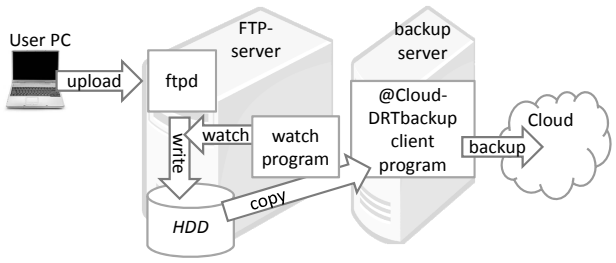


Figure 4. FTP server automatic backup system

added or updated file. This process consumes significant CPU resources, so we used the log file of the ftp daemon. The log file is updated after every transfer, and it contains all the information about added or updated or deleted files. When the watch program detects a modification of the log file, the updated files are copied from the FTP server to the backup server. In the backup server, the client program of "@Cloud-DRT backup" receives and backs up the updated files to the cloud.

In our prototype backup-system, the speed of performance of the file backup is only 1.4Mbps. The reason for this rather low performance is due to the use of the client program of "@Cloud-DRT backup" without modification. In this prototype, we introduced the personal use client program on the backup server. It will be necessary to build a new client program for an enterprise use as the next step.

B. Practical disaster recovery system demonstrating the advantage of mutually independent geographical sites

The essential configuration which has been developed for the data backup system which makes use of cloud environments is shown in Figure 5. The Meta-keys Processor uses several cloud environments after user authentication, and the multiple encrypted/divided/replicated data blocks are deployed in the different cloud environments, as judged appropriate, from the available network and computer resources.

In this section, we describe an equipment we have newly developed, which is effective not only for cloud systems, but also for on-premise systems. The equipment, named "DRTbox", is a small box with an ARM-CPU and Linux based OS. The DRTbox [13] is equipped with one Ethernet port and one serial port. It works under temperatures of 0 to 55 degrees centigrade and the power consumption is only

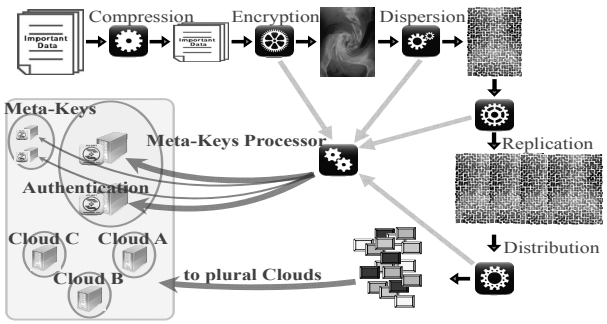


Figure 5. Data backup configuration using several clouds

5.0W. One of the typical configurations of the system is illustrated in Figure 6, which shows the case of a configuration in which a user site is backed up by two other sites. Data1 is the original data which is stored by a user. Data1a is stored in Site A, and Data1b is stored in Site B. Data1a and Data1b include HS-DRT-processed 32 blocks comprised of 64 blocks of whole data.

This means that even if one cloud out of three clouds does not work, the corresponding data will not be lost. Unlike conventional cloud systems, each site has its own file server for users. Let us consider the case where the data are processed by HS-DRT, and the size of Data1 is 100M bytes. Data1a and Data1b are each 50Mbytes in size. In this case, each site is basically controlled by a single company or an organization. Each site can be linked via a secure network. We used the technology to store the metadata keys for the other sites using a simple AES encryption. This makes the system operation simple, since the method to back up the metadata keys is complex and it is generally difficult to transport them securely.

If the user can use one more site (Site C), Data1 is divided into 64 blocks, each of 100/64 Mbytes, and processed to give 96 blocks. The additional 32 blocks are the parity data of the other 64 blocks. After that, the 96 blocks of data is divided into 3 groups, which are stored in Site A, Site B and Site C, respectively. This means that even if two sites out of four sites do not work, the corresponding data will not be lost. This configuration of the system is more robust than the one with just two back-up sites.

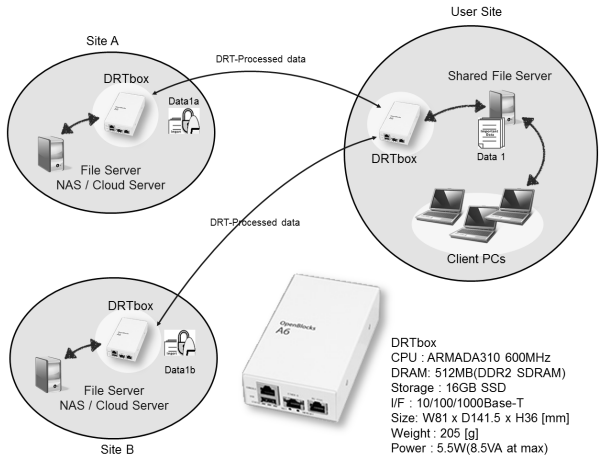


Figure 6. Implemented system using DRTBox

VI. INSIDER ATTACK ON BLOCK CIPHERS

A. Background

We need to consider case of an insider attack in which a malicious vender of a security system attacks his clients by using his cipher program, which is generally a block cipher. It can be shown that the Initial Value (IV) mode [11][12] of the block cipher is vulnerable to an insider attack. Moreover, we will propose a countermeasure to this attack.

While we are planning to provide a safe and secure system for the data backup, the client may be uneasy that the data might be stolen through an unexpected insider attack. To prevent this situation, we need to introduce the original block cipher mode, only for this purpose.

B. Newly occurring model of attack on block ciphers

In an attack on the block cipher based on the IV mode, the vendor as a supplier of the cipher program is assumed not to have any malicious intentions towards clients. The safety of block ciphers like the Advanced Encryption Standard (AES) has been discussed under such assumptions.

Cipher Block Chaining (CBC) mode or counter mode are recommended as safe methods to take the place of Electronic Code Book (ECB) mode. Since AES actually exhibits steady performance and reliability, it is now adopted as a world standard. However, one may often have a sub-conscious motivation for an insider attack. What sort of program could a vendor produce if the malevolent intentions and which might succeed in an insider attack? Such a situation raises a serious concern in the disaster recovery service environment which makes full use of cloud computing systems. We need to resolve this kind of problem when we offer disaster recovery services.

C. The model of an insider attack against block ciphers

First of all, let us define the model of an insider attack on a block cipher.

1) The vendor of the cipher program attacks the user of that program.

2) The cipher program outputs only the cipher text through the client user's input of key and message data. In particular, the cipher program is inaccessible to the network.

3) The user can preserve the cipher text, and he can always confirm the content by decoding according to the publicized coding method.

4) The attacker can obtain only the cipher text.

The IV mode of the block cipher is not safe, as follows. The attacker sends the client user encryption and decoding programs, called E(K,M), and D(K,C), respectively, as described below. Here K is a private key that only the client user knows, and M is a message.

A pseudo-random number IV adheres to the head of the cipher text C_0, such that $C=E(K,M) = (IV,C_0)$.

Here C is a cipher text that looks quite normal, and M can be computed by the normal decoding program D(K,C). However, the attacker can obtain the key K and so succeeds in the attack.

E(K,M) is composed as follows.

1. The attacker prepares internal code e and d in E(K, M).
2. Here, $e(M) = e(K_0, M) = C$, $d(C) = d(K_0, C) = M$. K_0 is a key that only the attacker knows.
3. The program E(K, M) receives the user's key K and plaintext M.
4. Let IV = e(K).
5. Calculate C_0 in a usual IV mode.
6. Output $C = (IV, C_0)$.
7. The attacker acquires C from the network.
8. The attacker can calculate K = d(IV).

When this method is implemented with hardware, it is especially difficult detect such an attack. This type of attack is not addressed by protocol analyzers such as AVISPA or Scyther since in their treatment of cryptographic primitives they adopt the so-called black box approach [14]. Even if we could detect the insider attack by using a protocol analyzer by reverse engineering a cryptographic primitive, we could hardly identify the same attack for another primitive. But, we propose an integration mode of block ciphers, as described below, which mode can avoid all insider attacks.

D. Vulnerability of double encryption method of block ciphers

Although the IV mode is not safe, it might be thought that it is safe if it is encrypted twice, including the appended pseudo-random number. However, it is understood that it is not safe if the attacker uses the following approach.

If $IV=D(K,IV_0)$, then , $E_0(K,C)=E_0(K,(IV,C_0))$

$=E_0(K,(D(K,IV_0),C_0))=(IV_0,C_1)$, where E_0 is the corresponding ECB mode encryption .

Hence, it is assumed $IV_0=e(K)$ by using a secret internal code e that only the attacker knows.

E. Integration mode of block cipher

Now, we propose an encryption mode that is not vulnerable to an insider attack. This encryption mode does not have redundancy, and has the specific characteristic of not being vulnerable to an insider attack. Moreover, it has security more than equal to that of the block cipher which was originally used.

We can provide a proof for this using a non-trivial one to one correspondence

$f:M \rightarrow f(M)$ (integration transform).

Let E(K,M) and D(K,C) be the encryption and decryption, respectively, of ECB mode of a safe block cipher. In E(K,M) we adjust the length of M by zero padding. At this time, the encryption and decryption are defined by $C = E(K, f(M_0))$ and $f^{-1}(D(K, C))$, respectively, where M_0 denotes the zero padding of M. In what follows, let Length(M) be the number of bits of any message M.

Theorem 1. If the block cipher $E(K, M_0)$ is safe then $E(K, f(M_0))$ is also safe.

(Proof) For some algorithm A, suppose $A(E(K, f(M_0))) = M_0$, then we have

$f(A(E(K, M_0))) = f(A(E(K, f(f^{-1}(M_0))))) = f(f^{-1}(M_0)) = M_0$.

Therefore the original block cipher E is also unsafe.

Theorem 2. $E(K,f(M_0))$ is safe against an insider attack.

(Proof) Let $C = E(K,f(M_0))$ and we assume this is not safe against insider attacks. Note $Length(M_0) = Length(C)$. Since $Length(K) > 0$, we have $Length(M_0) + Length(K) > Length(C)$ and note that C contains the information of K and M_0. Therefore, the attacker must use a compression algorithm, so we cannot decrypt C by the compatible decryption program. This is a contradiction.

As mentioned above, client users of the disaster recovery system are generally vulnerable to potential insider attack by the system vendor.

 ISBN: 978-1-61208-305-6

However, by making use of the method proposed here, this kind of vulnerability can also be avoided by using the integration mode of AES or ECB mode.

VII. CONCLUSION AND FUTURE ISSUES

We have presented several types of commercialized system and performance results for a system using HS-DRT in cloud computing environments.

Further studies should address the optimum network utilization technology. We are planning to verify the essential characteristics necessary to fully utilize network resources, in order to commercialize an ideal disaster recovery system. In the conventional disaster recovery system mentioned above, we assumed the use of push-type client nodes which have ID information registered in advance in both the data center and the supervisory center. However, it would be preferable to increase the number of potential users to support the corresponding network services. For this purpose, client nodes such as those in a mobile cell-phone network can be utilized for the network services, even if their IP addresses are changeable with time. To effectively realize this scheme both the data center and the supervisory center register each list of fragmented file information relating to each client node ID with a changeable IP address, in preparation for a recovery request from the user. We should further examine the performance of the Web server as the number of fragments increases.

Since the formal protocol analyzers cannot include the reverse engineering for all the cryptographic primitives, they cannot correspond to all the proposed insider attacks.

In contrast, we have proposed the integration mode of block ciphers, and this provides verifiable security against all the proposed insider attacks.

ACKNOWLEDGMENT

This work has been partially supported by the study (Issue number:151) of the National Institute of Information and Communications Technology (NICT) of Japan.

REFERENCES

[1] N. Miyaho, Y. Ueno, S. Suzuki, K. Mori, and K. Ichihara, "Study on a Disaster Recovery Network Mechanism by Using Widely Distributed Client Nodes," ICSNC 2009, pp. 217-223, Sep., 2009.

[2] Y. Ueno, N. Miyaho, S. Suzuki, and K. Ichihara, "Performance Evaluation of a Disaster Recovery System and Practical Network System Applications," ICSNC 2010, pp. 195-200, Aug., 2010.

[3] S. Suzuki, "Additive cryptosystem and World Wide master key," IEICE technical report ISEC 101(403), pp. 39-46, Nov., 2001.

[4] N. Miyaho, S. Suzuki, Y. Ueno, A. Takubo, Y. Wada, and R. Shibata, "Disaster recovery equipments, programs, and system," Patent. publication 2007/3/6 (Japan), PCT Patent :No.4296304, Apr., 2009.

[5] K. Kokubun, Y. Kawai, Y. Ueno, S. Suzuki, and N. Miyaho, "Performance evaluation of Disaster Recovery System using Grid Computing technology," IEICE Technical Report 107(403), pp. 1-6, Dec., 2007.

[6] S. Kurokawa, Y. Iwaki, and N. Miyaho, "Study on the distributed data sharing mechanism with a mutual authentication and meta-database technology," APCC 2007, pp. 215-218, Oct., 2007.

[7] S. Tezuka, R. Uda, A. Inoue, and Y. Matsushita, "A Secure Virtual File Server with P2P Connection to a Large-Scale Network," IASTED International Conference NCS2006, pp. 310-315, Mar., 2006.

[8] R. Uda, A. Inoue, M. Ito, S. Ichimura, K. Tago, and T. Hoshi, "Development of file distributed back up system," Tokyo University of Technology, Technical Report, No.3, pp. 31-38, Mar., 2008.

[9] Y. Deswarte, L. Blain, and J. C. Fabre, "Intrusion tolerance in distributed computing systems,"Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on , pp. 110-121, May, 1991

[10] R. A. Fisher and F. Yates, Statistical tables for biological, agricultural and medical research (3rd ed.). London: Oliver & Boyd. pp. 26–27, 1948.

[11] J. Katz and Y. Lindell, "Introduction to modern cryptography", Principles and Protocols, Chapman & Hall/CRC, pp. 96-106, 2008.

[12] D.R. Stinson, "Cryptography, -Theory and Practice- 3rd edition",Chapman & Hall/CRC, pp. 73-112, 2006.

[13] N.Nojima, "The DRTbox", pp. 1-2, Aug., 2013. http://www.nogic.net/files/AtCloudDRTbox.pdf.

[14] C.J.F. Cremers, "Scyther-Semantics and Verification of Security Protocols", Ph.D. Thesis, Eindhoven University of Technology, pp.11-12, 2006, ISBN 90-386-0804-7. ISBN 978-90-386-0804-4.

[15] R.J. Anderson, "The eternity service", Jun., 1997. http://www.cl.cam.ac.uk/~rja14/eternity/eternity.html, 2013.8.22.

 ISBN: 978-1-61208-305-6

Identity Management Approach for Software as a Service

Georgiana Mateescu, Marius Vlădescu

Computer Science and Automatic Control Faculty
Polytechnic University of Bucharest,
Bucharest, Romania
georgiana.mateescu@gmail.com, vladescumariusnicolae@yahoo.com

Abstract— **Cloud Computing is defined by flexibility, agility, scalability, reliability – all these being provided using the concept of computing as a utility. Beside all its big advantages, the adoption of the cloud still has a limited number of adherences because it also can expose the consumer to a lot of risks and vulnerabilities, which sometimes are heavier than the benefits themselves. Providing a cloud identity is a key component of a cloud successful story because in such architecture aspects like the request context, sensitive data usage and end-user service availability make the difference between a legitimate and an un-legitimate request. By using several systems that provide only parts of the cloud identity, we ensured that the end user cannot alter his access to cloud application and that the cloud application only manipulate allowed personal data.**

Keywords-Cloud Computing; identity management; Software as a Service.

I. INTRODUCTION

Cloud computing, as defined by National Institute of Standards and Technology (NIST), is a model for enabling always-on, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., storage, applications, services, etc.) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1].

This definition provides the main characteristics of cloud computing [2]:

- Shared resources – in cloud computing architectures multiple users utilize the same resources from network level, host level to application level.
- Massive scalability – because of its foundation principles, cloud computing has the ability to scale to thousands of systems.
- Elasticity – in cloud computing framework it is very easy to adapt both hardware and software resources to the user's need.
- Pay as you go – cloud computing consumers pay only the resources they use just for the time they actually require them.
- Self provisioning of resources – additional systems (processing capability, software, and storage) and network resources are added when and if they are needed.

These characteristics prove a lot of advantages including: lower-cost computers for users, improved performance, lower IT infrastructure costs for enterprises, fewer maintenance issues, lower software costs, instant software updates, increased computing power, unlimited storage capacity, increased data safety, improved compatibility between operating systems, improved document format compatibility, and universal access to documents [3]. There is also a list of challenges that can be even heavier than the benefits:

- This type of architecture requires constant Internet connectivity; moreover the disconnection can lead to a lot of inconsistencies that can be very hard to clean.
- The services quality is dependent on the connectivity power; it does not matter how strong and reliable a service is; if for bad weather for example, the Internet is slower, the entire performance is compromised
- The data ownership is shared between cloud provider and consumer, which may lead to insufficient security in storing and using the data by the cloud vendor.
- Compatibility issues between the cloud providers can lead to delays in processing data from systems hosted by different vendors, inefficiency in meeting the required Service Level Agreements (SLAs).
- Regulatory compliance issues due to the geographical position of the cloud provider data center.

In order to phase all these challenges, the community proposed standards which came in the help of enterprises and offer them guidelines about what secure architecture means and. Also, from decades of experience, the Internet brought out to light a lot of best practices related to the information security, vulnerabilities, risk assessment and damage management and recovery. In this paper, we will reference the Cloud Security Alliance (CSA) standard that addresses the main security topics within cloud architectures [4].

The first section of this paper presents the actual IT context for cloud computing concept with both its advantages and disadvantages. The second section states the deployment and services models and after this the main security domains together with the challenges from a cloud computing architecture are presented. Identity management within a hybrid company (it uses both on premises applications and on demand services) is the key component that can make the difference between a successful cloud computing story and a failed one. In the fourth section, we propose an approach that addresses the identity management requirements by leveraging the existing architecture within the company. In the last section, before referencing all state-of-the-art cloud computing materials, we conclude with benefits and future work.

 ISBN: 978-1-61208-305-6

II. CLOUD COMPUTING DEPLOYMENT AND SERVICE MODELS

According to Lingenfelter [4], cloud computing is defined by the following characteristics:

- On-demand tenant self-service model for provisioning computing capabilities
- Broad network access and mobile platforms
- Resource pooling through dynamically assigned physical and virtual capabilities delivered in a multi-tenant model and location independent
- Rapid elasticity of provisioned resources
- Measured service to monitor, control and report on transparent resource optimization
- All these features can be delivered as one of the three types of services:
- Software as a Service
- Platform as a Service
- Infrastructure as a Service / Datacenter as a Service

The cloud computing benefits can be implemented using one of the four deployment models depicted in Figure 1.

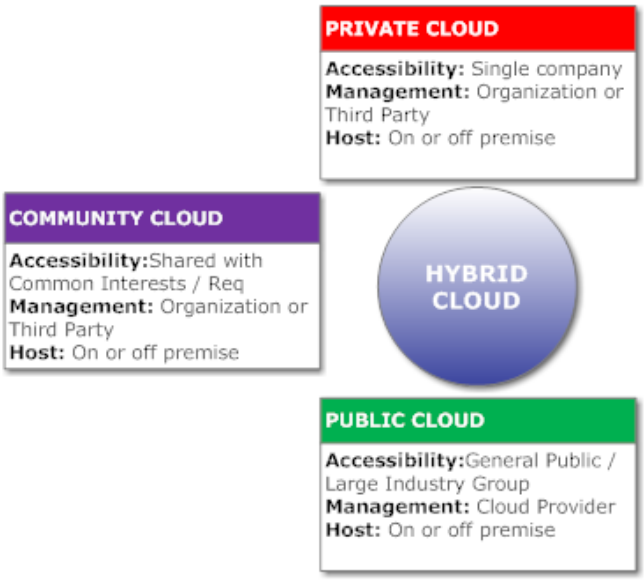


Figure 1. Cloud Computing Deployement Models [13]

The private cloud option is ussualy adopted by companies with multiple braches that choose to maintain all the applications within the organization's firewall having all the benefits the cloud architecture provides.

The community cloud offers the cloud benefits at a community level – a community is a group of organizations that address the same domain of activities. In this model, the cloud services are not kept within firewalls, but are restricted based on community membership.

The public cloud is the least restrictive from access perpective and it has a general use.

Each of these models has certain challenges that lead to the hobrid approch – the combination between private clouds (which usually hosts the most critical application), community clouds (which hosts the applications required for organizations interoperability in every day business process) and public clouds (which hosts applications for general usage – such as customer portals etc.).

Depending on each specific environment and business requirements, the cloud model should be selected after a complex process that involves both technical and functional (business) teams.

III. CLOUD COMPUTING SECURITY CHALLENGES

European Network and Information Security Agency (ENISA) provided a detailed description of all risks that a cloud customer must phase together with the impact and affected areas specific on each risk [2].

When adopting a cloud computing architecture, the company must address a significant number of vulnerabilities and risks including [2]:

- Policy and organizational risks include: Lock-in, Loss of governance, Compliance challenges, Loss of business reputation due to co-tenant activities, Cloud service termination failure
- Technical risks include: resources exhaustion, isolation failure, cloud provider malicious insider – abuse of high privilege role, intercepting data in transit and data leakage on up/download intra-cloud, loss of encryption keys, Economic/Distributed denial of service (EDoS/DDoS)
- Legal risks
- Risks not specific to the cloud include: network management (breaks, congestions, miss-connection, non-optimal use), privilege escalation, backup lost, stolen, unauthorized access to premise, natural disaster
- Vulnerabilities include: AAA vulnerabilities, user provisioning and de-provisioning, remote access to management interface, lack of resource isolation (resources that are used by one customer can affect resources used by another customer), lack of reputational isolation, communication encryption vulnerabilities

In order to efficiently address all these risks and vulnerabilities, CSA categorized all the security aspects from cloud computing architectures into the following domains[4]:

1. Governance and Enterprise Risk Management - The ability to govern and measure risk introduced by cloud computing
2. Legal Issues: Contracts and Electronic Discovery - Security breach disclosure law
3. Compliance and Audit - Evaluate how cloud affects compliance
4. Information Management and Data Security - Managing data stored in cloud
5. Portability and Interoperability - The ability to move data from a cloud provider to another
6. Traditional Security, Business Continuity and Disaster Recovery - How cloud affects the current security procedures

 ISBN: 978-1-61208-305-6

7. Data Center Operations - How to evaluate provider's data center architecture and operations
8. Incident Response, Notification and Remediation - Proper incident detection, response, notification and remediation
9. Application Security - Securing application that runs on different cloud deployment model
10. Encryption and Key Management - Identify proper key usage and key management
11. Identity and Access Management - Cloud-based IdEA (Identity, Entitlement and Access Management)
12. Virtualization - Risk associated with VM isolation, VM co-residence
13. Security as a Service - Third party security assurance including incident management and compliance attestation

In this paper, we will address the Identity and Access Management domain and we will present our approach that leverages the existing Identity Manager system within the company premises in order to accommodate all the cloud specific requirements.

IV. IDENTITY AND ACCESS MANAGEMANT

A. Existing identity management approaches

An Identity Management component is in charge with the following tasks:

- Establish identities: Associate personally identifiable information with an entity.
- Describe identities: Assign attributes identifying an entity.
- Record the use of identity data: Log identity activity in a system and/or provides access to the logs.
- Destroy an identity: Assign expiration date to personally identifiable information. Personally identifiable information becomes unusable after the expiration date.

Identity Management can involve three perspectives [11]:
1. The pure identity paradigm: creation, management and deletion of identities without regard to access or entitlements.
2. The user access (log on) paradigm: A traditional method say for ex a user uses the smart card to log on to a service.
3. The service paradigm: A system that delivers personalized role based, online, on-demand, presence based services to users and their devices.

P. Angin [7] provides an approach to create and manage identities within cloud computing architectures without the need of trusted sources. This approach does not use enterprise systems and it is based on the information stored locally on the machine that attempts to connect to the cloud service.

C. Mitchell [8] describes an identity management mechanism that performs privacy-preserving authentication using anonymous credentials without making usage of any enterprise system.

Waleed Alrodhan [9] provides a browser plug-in approach for user authentication. Every digital identity is a security token. A security token consists of a set of characteristics, such as a username, user's full name, address, SSN etc. The tokens prove that the claims belong to the user who is presenting them. The biggest challenge related to this approach is related to the user behavior: they do not understand the importance of the approval decision or because they know that they must approve the certificate in order to get access to a particular website.

P. Mularien [10] provides a decentralized authentication protocol that helps cloud users to manage their multiple digital identities by providing one set username and password—an OpenID which is further used for cloud authentication. The main disadvantage of this approach is its susceptibility to phishing attacks and social engineering.

Each of the above solutions has its own challenges and risks.

B. Personal approach

In our scenario, we have the following systems:

- On premise Identity Manager system – this system stores all the company employees together with their accounts in the enterprise systems and the access policies that define the systems and services that the users are allowed to use. In order to efficiently implement the access policies, these are configured based on specific employees attributes also stored in the Identity Manager solution. Also this application is the cloud identity issuer that generates the cloud identities based on different components provided by the other applications involved in the authentication process.
- On demand Trusted Certificate store – this system provides digital certificates to the company. Its main task is to verify the digital certificate from the validity, legitimacy and data integrity perspectives.
- On demand cloud services – we used 2 services: Business Intelligence (BI) application and a Customer Relationship Management (CRM) solution

Within a cloud computing architecture, various actors from both on demand and on premises systems are involved in the authentication process. These parties are [6]:
1. Identity Provider (IdP) – this component issues digital identities. In our scenario, the identity provider in the on-premise Identity Manager system and it is a trusted identity store for the cloud services.
2. Service Provider (SP): It provides access to services to the identities that have the right required identities. Our use-case service model in Software-as-a-Service – more specifically we used a Business Intelligence (BI) application and a Customer Relationship Management (CRM) solution.
3. Entity: Entities are the ones about who claims are made. In our scenario, we consider the entity the user himself who wants to access one of the available cloud services.
4. Identity Verifier: Service Providers send them the request for verifying claims about an identity. In our use-case the identity verification is performed using Trusted Certificate Store and it has the particularity that this step is performed by the Identity Provider itself.

 ISBN: 978-1-61208-305-6

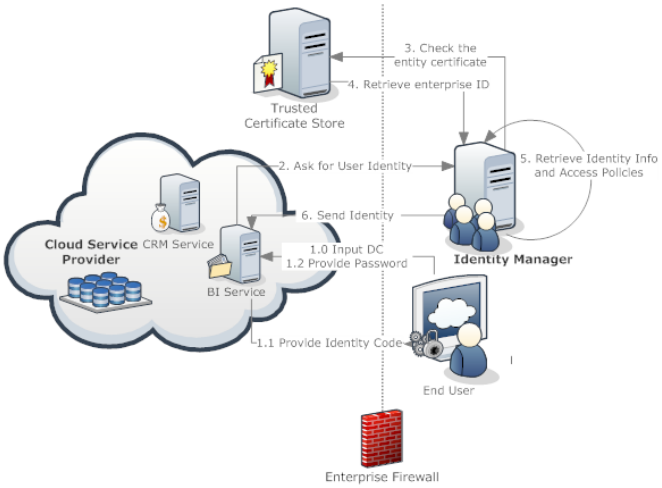


Figure 2. Identity Management Flow

Figure 2 depicts the identity related information flow:

- The user has a token provided by the cloud service provider that generates the password for the cloud service based on a challenge code received from the service provider.
- On the enterprise machine (the user's computer) it is stored a Digital Certificate (DC) that contains, beside the certificate's identifications, the user's enterprise unique identifier and also information related to the device from where the certificate is being used to perform authentication.

The authentication process has the following steps:

1. The user accesses the cloud service and in order to authenticate, he issues the digital certificate.
2. The cloud service answers with a code used by user's personal token to generate the password. The user introduces then the generated password in the service login window.
3. If the password is correct, then the cloud service sends the DC to Identity Manager together with the entity code (the one generated in step 2).
4. Identity Manager sends the DC to the Trusted Certification Store in order to ensure that the certificate is valid and legitimate and to verify data integrity. A certificate is considered *valid* if its expiration date is greater than the day when this authentication request that uses the certification is performed. A certificate is considered *legitimate* if the authentication request that used it, was raised from a host that is allowed to use that certificate. The data integrity is achieved when the personal information of the user stored on the certificate is the same as the one from the store. If all three validations are successfully completed, the Trusted Certificate Store retrieves the enterprise unique id to Identity Manager.

All the data within the digital certificate are encrypted.

5. Using the enterprise unique id, Identity Manager generates the cloud identity based on the specific employee attributes.
6. The Identity Manager encrypts the cloud identity using the key received from the cloud service and sends it back to the service.
7. Based on the allowed accesses for that specific identity, the available services are displayed to the client.
8. After all the activities performed in the cloud service are completed, when the user logs out, the cloud identity is automatically destroyed by the cloud provider.

If any of the validation described in the process fails, then an error is displayed to the user and the access to the cloud services is not granted.

The above presented flow is described from the different systems interactions point of view in Figure 3.

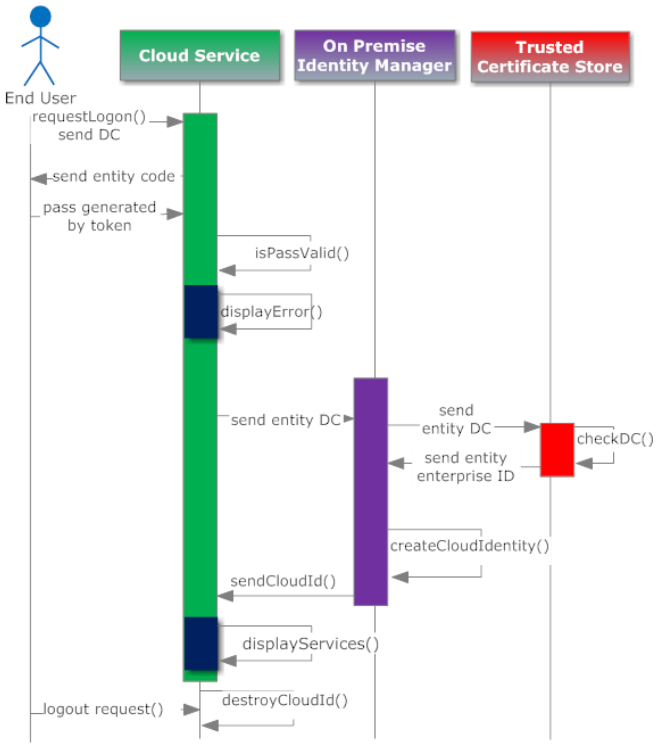


Figure 3. Authentication Process

The cloud identity generated by the Identity Manager system from the enterprise premises contains the following information (Figure 4):

- Personal information about the entity required in order to perform specific tasks in the cloud service that the user is allowed to use (such as fist and last name, personal account, social security number etc).
- Data related rights required to define which operations from the cloud service can use specific cloud identity information (such as the only the payment service can read the account data).
- Access policies required to define which operations from the cloud service are allowed to the user (such as if the user is just a regular user, he is only allowed to purchase some product or to see only his previous transactions).

 ISBN: 978-1-61208-305-6

- Information related to the cloud identify sender – this is used in order to protect against man in the middle attacks, although we considered the communication between the Identity Manager and the Cloud Service trusted (such as the physical address of the computer that made the request to access the cloud service).

In order to protect identity cloud data integrity, we used a symmetric encryption algorithm [12] (the encryption key is the code generated by the cloud service provider in the first step of the authentication).

Figure 4. Cloud Identity

Figure 5 depicts the data usage of the cloud identity within the cloud service:

- The cloud service (Business Intelligence application) decrypts the cloud identity information
- The cloud service (Business Intelligence application) reads from the access policies the services that must be provided to the user
- When the user performs a certain operation, the Business Intelligence application access the personal information according to the data rights from the cloud computing.

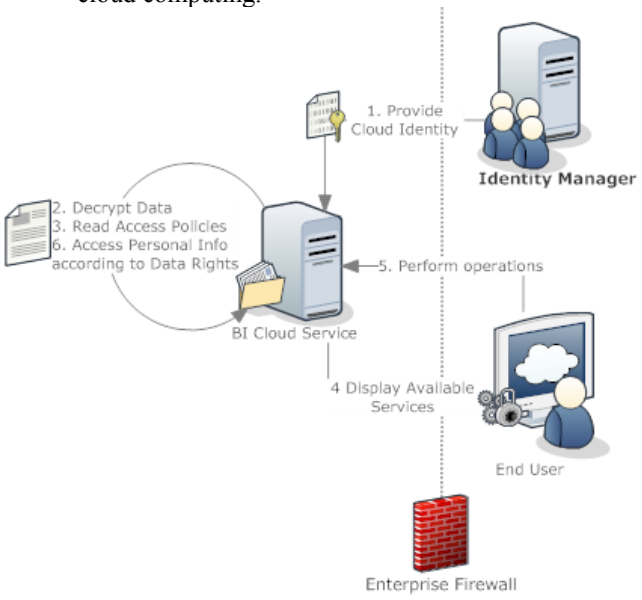


Figure 5. Cloud Identity usage

By using this approach, we managed to control the cloud application user access and also to protect the user data from being used in an abusive manner. The cloud identity ensures that the cloud application will use only the data it is allowed to and that after the access to the cloud application is finished the cloud identity is destroyed.

V. CONCLUSION

According to Jansen and Grance [6], cloud computing guarantees for enhanced performance with lower cost; but, in the same time, the adopters must balance the cost reduction with the security and privacy measures.

In this paper, we presented an identity management approach that allows the enterprises to provide a secure transition step – from on premise architecture to on demand on. This transition supposes a hybrid architecture containing both on premise and on demand systems that interact between each other in everyday business flows.

Providing the fact that on-demand services represent an extension of on-premise applications, existing security techniques can be applied within individual components of cloud computing. We used the on-premise identity provider system to generate cloud identities. This was possible because the identity manager not only stores the identities themselves, but it also has validation and verification capabilities and stores all the policies required to properly set the access to the user.

The main benefits of our approach are:

- The user does not know at any time his cloud identity, thus he will never be able to alter his access rights into the cloud systems.
- The two steps verification process protects against the theft of digital certificate:
  - The first step is the password generated by the token, using the code from the cloud provider – if the attacker steals the DC without having the token, he will not be able to pass the first authentication step
  - The second verification is performed by the Trusted Certificate Store that checks if the certificate issuer is legitimate – this means that even if the attacker managed to steal also the token from the user, if he uses the certificate from another device the authentication process will fail.
- The authentication process involves multiple systems, which means that the man in the middle attack is impossible to be efficiently exploited with a single penetration step.
- The cloud identity contains only the required information according to specific access policies and also data access rights within the allowed services.
- The cloud identity is destroyed after usage, without storing it in the cloud.

Compared to similar papers listed in the beginning of Section IV, we used multiple systems is the identity generation process, fact that minimizes the risk of attacking the identity issuer system. Also, we encrypted the cloud identity data in order to ensure that if it is captured after the entire generation

 ISBN: 978-1-61208-305-6

process is completed, the attacker will not be able to access the data stored in it.

For future works, we plan to address two more areas within the proposed approach:

- The implementation of a more sophisticated encryption algorithm for the cloud identity data
- The single sign on process within on premises and on demand systems.

REFERENCES

[1] Timothy Grance and Peter Mell, "The NIST Definition of Cloud Computing", National Institute of Standards and Technology, U.S. Department of Commerce, Special Publication 800-145, Sept 2011.

[2] Daniele Catteddu and Giles Hogben, "Cloud Computing Security Risk Assessment", European Network and Information Security Agency (ENISA), 20 Nov. 2009

[3] Jim Reavis , "Security Guidance for Critical Areas of Focus in Cloud Computing v3.0", 14 Nov. 2011

[4] David Lingenfelter, "Cloud Audit and Cloud Trust Protocol", 2011, unpublished, last accessed on October, 5th,2013. https://cloudsecurityalliance.org/research/grc-stack/

[5] Tim Mather, Subra Kumaraswany, and Shahed Latif, "Cloud Security and Privacy. An Enterprise Perspective on Risk and Compliance", O'Reilly United States of America, ISBN-10: 0596802765, 5 Oct. 2009

[6] Wayne Jansen and Timothy Grance , "Guidelines on Security and Privacy in Public Cloud Computing", U.S. Department of Commerce, Special Publication 800-144, December 2011

[7] Pelin Angin, Bharat Bhargava, Rohit Ranchal, Noopur Singh, Lotfi Ben Othmane, Leszek Lilien, and Mark Linderman , "An Entity-centric Approach for Privacy and Identity Management in Cloud Computing", Reliable Distributed Systems, 2010 29th IEEE Symposium, 31 October 2010

[8] Simone Fischer-Hübner, Elisabeth de Leeuw, and Chris Mitchell : "Policies and Research in Identity Management" - Third IFIP WG 11.6 Working Conference, IDMAN 2013, London, UK, April 8-9, 2013, ISBN 978-3-642-37281-0

[9] Waleed A. Alrodhan and Chris J. Mitchell "Improving the Security of CardSpace", EURASIP Journal on Info Security Vol. 2009.

[10] Peter Mularien, "Opening up to OpenID with Spring Security", May 2010.

[11] Sandeep K. Sood, Anil K. Sarje, and Kuldip Singh, "A secure dynamic identity based authentication protocol for multi-server architecture", Journal of Network and Computer Applications, Volume 34, Issue 2, March 2011, pp. 609-618

[12] Masashi Une and Masayuki Kanda, "Year 2010 Issues on Cryptographic Algorithms", Discussion Paper No. 2006-E-8, IMES, C.P.O BOX 203 Tokyo, 100-8630 Japan

[13] Neeraj Metha, "The 4 Primary Cloud Deployment Models", CloudTweaks, July, 2nd, 2012

 ISBN: 978-1-61208-305-6

OTIP: One Time IP Address

Renzo Davoli
Computer Science and Engineering Department
University of Bologna
Bologna, Italy
Email: renzo.davoli@unibo.it

***Abstract*—One Time IP address (OTIP) is a security feature to protect private communications on the Internet. OTIP enables nodes to change their IP addresses periodically following a cryptographic sequence. Legitimate users have all the information needed to compute the current addresses used by the servers, thus their networking clients are able to access the required services. OTIP current address is the output of a computation based on the Fully Qualified Domain Name (FQDN) of the server, a secret password and the current time. The major achievement of OTIP is that all the IP addresses collected by wiretapping the networks are useless for attackers as, in a short time, all the servers will be using different addresses.**

***Index Terms*—IP networks; TCPIP; Information security.**

I. Introduction

OTIP means that the current IP address of a server changes periodically to prevent networking attacks. The current IP address of a server is computed on the basis of some private information shared by legitimate users and the server itself, like a password, and the current time. This method has mainly been designed for IPv6 networks. In fact, the current server address can be picked up as one of the valid host addresses available on the local network, most of the time among 2^{64} possible addresses. Clearly, a 2^{64} address space is too large for attackers to try a brute force enumeration attack on all the available addresses; even if they eventually succeeded, the retrieved addresses would have to be exploited before their validity expires and the servers move to new addresses.

In theory, the same method could also be applied to IPv4, but it would be ineffective, due to scarcity of addresses and to the narrowness of address spaces within a network, most of the times 256 nodes or less.

The paper is organized as follows: the next section discusses the proposal, while section III compares OTIP with other methods known in the literature. Section IV, which gives an estimation of the probability that two servers may temporarily choose the same address, is followed by the implementation section, which describes the experimental part of the paper. The final section is about the limits, future developments and conclusive remarks.

II. Discussion

The Internet supports services for the general public as well as private services for a predefined set of authorized users. For a business firm or a company, its institutional web site is generally a service provided for the general public. On the other hand, a remote shell service for system administration or a Voice over IP (VOIP) service interconnecting the company's (software) Private Branch Exchanges (PBX) are examples of private services.

All the private servers clearly have a means of protection to prevent access by unauthorized users (e.g., password protections and traffic encryption). OTIP aims to provide one further layer of protection for private services.

Without OTIP, attackers can collect the IP addresses of the servers by wiretapping the network and creating a catalog of valid addresses and services. These addresses can later be used to perform brute force attacks, for instance using a database of weak passwords, or even to test vulnerability by using a collection of well-known exploits of the servers' code.

OTIP can prevent these attacks, or at least makes them extremely hard to succeed, as the addresses collected by network sniffers expire in a short time.

OTIP does require the Real Time Clocks (RTC) of servers and clients to be synchronized, as the current time is a parameter to compute the current address. Networking services for RTC synchronization, like NTP (Network Time Protocol) [1], are quite common. Modern NTP implementations use authenticated servers, which append a confirmation magic number to prove the authenticity of each synchronization packet. The use of authenticated NTP servers is warmly suggested for OTIP. Although misaligned clocks would not allow the discovery of the current IP addresses for servers, attackers could tweak the clients' perception of the current time, thus preventing the legitimate users from accessing their services. In other words, an attack on NTP would cause a DoS (Denial of Service) for OTIP.

NTP, or other RTC synchronization protocols, is able to reduce the error between the time read at different hosts below a certain predefined level. These protocols, in fact, periodically check the time of the current host against the time provided by reliable synchronization servers on the network. The difference between the local time and the time retrieved from the network corrected using an estimation of the communication delay, is used to put in place some modifying actions on the local RTC current value (e.g., by tuning the frequency of the local clock).

Even when the value of the current time can be retrieved from a very precise service, all the actions do not take place on clients and servers simultaneously, due to the transmission delays of the network. For example, a client's request takes time to travel across the network. Thus, the time read by the client when the request is issued will never be the same of the

 ISBN: 978-1-61208-305-6

time at which the server receives and processes that request.

In order to deal with all these errors and delays OTIP cannot manage the address change as an instantaneous action. For a specific time interval both addresses (the old one and the new one) should be valid. This time interval should be carefully chosen to solve two possible problems.

- A request from a client whose RTC is running slightly fast (while remaining within the admitted tolerance) may arrive in advance with respect to the time due for the address change on the server.
- A request from a client whose RTC is precise or slightly late can arrive at the server after the time due for address change, also because of the delay introduced by the networking communication.

In order for a server to tolerate clock desynchronization and network delays, the new address should be activated in advance for at least a time equal to the maximum difference between the RTC readings, and the old address should be kept valid for a time at least equal to the maximum transmission delay in the network plus the maximum difference of the RTC.

Formally, calling Δ_t the maximum desyncronization guaranteed by the clients and the server, i.e., if s is the server and c a generic client:

$$\forall c : |t_c - t_s| < \Delta_t \tag{1}$$

and when the maximum network delay is d_{net}, the period of the OTIP address change is T, then the n-th server address must be valid and available for clients in the time interval:

$$[t_0 + (n-1)T - \Delta_t, t_0 + nT + \Delta_t + d_{net}] \tag{2}$$

III. Related Work

A time based One Time Password (OTP) [2] is a password which is valid for a limited time. This security feature is commonly used to protect transactions on the network. Many Internet based banking systems, for instance, provide each user with their own small piece of hardware called *security token*. This device, usually similar to a key-holder, shows on its display a password generated on the basis of a secret seed and the current time provided by a reasonably precise clock which is a hardware component of the security token itself. OTIP applies the OTP concept to IP addresses instead of passwords.

IETF (Internet Engineering Task Force) have introduced by RFC3041 [3] and then by RFC4941 [4] the idea of dynamically changing IP addresses. The focus of these standards (and of other proposals like [5]) is the privacy of the client. Autoconfiguration methods, in fact, can reveal the MAC (medium access control) address of each host connected to the Internet because that address is copied in some bytes of the IPv6 address, allowing attackers to trace the position of a specific computer on the network. OTIP and these privacy extensions have different purposes: the former applies the dynamic change of IP addresses to protect server from attacks, the latters changes the IP addresses of the client to preserve the user's privacy.

OTIP is a *killer application* of the Internet of Threads (IoTh). As described in [6], IoTh opens a wide range of new applications by allowing each process to be a node of the Internet. Each process can have its own IP addresses, routing definitions etc.

Using IoTh each OTIP server can define its own OTIP policy, address change period, address computation algorithm, overlapping time interval for address validity etc.

The proof-of-concept implementation provided in section V, is based on LWIPv6 [7], View-OS [8] and msocket [9].

OTIP could be implemented without IoTh using a daemon to add and delete IP addresses of a network controller or a virtual interface of a container [10]. In both cases this daemon would be granted network administration capabilities (`CAP_NET_ADMIN` in Posix.1 [11] terminology). In this scenario the selection of the right address to use should be done by the `bind` system call. Other processes running on the same host may erroneously use the same address designed just for a specific server (e.g., by using `in6addr_any`, as many daemons do). Apart from the software architectural complexity of having a daemon as an executor of the OTIP address change requests, the effects of an attack would not be confined to a single service but could scale up to the container or to the whole networking support of the hosting system. IoTh implementation is clearer, simpler and safer. Finally, can be regarded as a special case of a hash-based address as defined in [12].

IV. Address Collision

It is clearly possible, although improbable, that two OTIP servers running on the same data-link network (real or virtual Local Area Network, LAN) temporarily get the same address. If we regard the addresses as if they were randomly chosen, this problem can be regarded as an application of the Birthday Problem (also known as the Birthday Paradox, as explained in [12]).

The probability of m nodes choosing the same address using a host suffix of h bits is:

$$Pr[(h,m)] = 1 - \frac{2^h! \binom{m}{2^h}}{m^{2^h}} \tag{3}$$

which can be approximated when $m << 2^h$:

$$Pr[(h,m)] \approx 1 - e^{-\frac{m^2}{2^{h+1}}} \tag{4}$$

Figure 1 shows the probability of address collision using a 64 bit network prefix and a 64 bit host address, which is the most common scenario in current IPv6 implementations.

The probability in a network connecting one thousand servers has the order of magnitude 10^{-14}, and even connecting one million servers the collision probability is less than 10^{-7}.

The effect of a collision is a temporary unreachability of the servers lasting for an OTIP address change period (64 seconds in the proof-of-concept implementation). Clearly this limitation should be taken into account for applications which require extremely stringent constraints in service continuity,

 ISBN: 978-1-61208-305-6

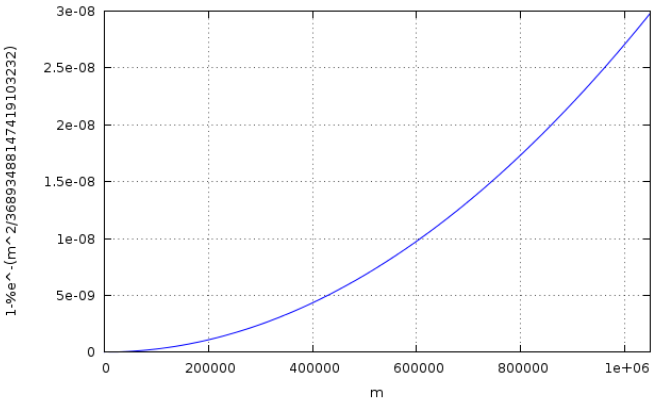


Fig. 1. Probability of address collision in a :64 IPv6 network [12]

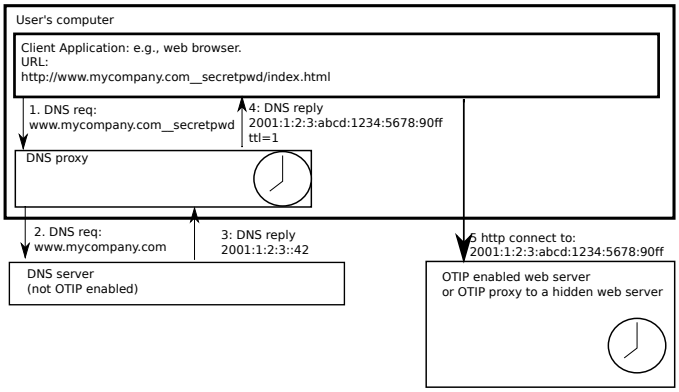


Fig. 2. Structure of the proof-of-concept implementation of OTIP for TCP communications

but the eventuality of service fault for address collision is able to satisfy the requirements of a very wide range of network services.

V. Implementation

This section describes a proof-of-concept implementation of OTIP for TCP (transmission control protocol) servers and some ideas for alternative implementations also for connection-less services based on UDP (user datagram protocol).

Figure 2 provides a schematic view of the test implementation. In order to support unmodified client programs the prototype uses an OTIP enabled Domain Name Server (DNS) proxy which is installed on the user's computer where the client program is running. This DNS proxy transparently forwards all the requests to the real DNS server except those matching the OTIP syntax. For the scope of this prototype the syntax is:

```
FQDN__password
```

So, if in a configuration file of the client program or, for instance, in the host part of a URL of a browser a specification like the following:

```
www.mycompany.com__secretpwd
```

appears, it will be resolved by the DNS proxy using the base address of `www.mycompany.com` retrieved by a query to the local DNS server (or possibly from the DNS server of `mycompany.com`). The host part of the resulting IPv6 address will be computed by the proxy using the base address, the password (`secretpwd` in this example) and the current time. Calling t the current system time in seconds (as returned by the `time` POSIX system call), the method implemented in this proof-of-concept computes a 128bit MD5 hash of the string composed by $t \gg 6$ (right shifted 6 bit positions, i.e., divided by 64), a space and the password. The statement used to create the input string for the hash function is the following:

```
len=asprintf(&s,"%d %s",time(&now)>>6,pwd);
```

The result of an exclusive-or (XOR) operation between three operands:

- the 64 most significant bits of the MD5 hash value,
- the 64 least significant bits of the same value,
- and the host part of the address returned by the real DNS,

becomes the host part of the current OTIP address. It is worth noting that the description of the algorithm used to compute the OTIP address in this implementation has been given here for the sake of presentation completeness. Any choice of hash function involving the time and the FQDN or IP address returned by the real DNS would fit, provided the same algorithm is consistently applied both by the server and the client.

The proxy sets the TTL (time-to-live) to one second in its reply to force the client to invalidate the resolution cache in a short time and to repeat the query for any further connection needed to the same server. This is necessary as the address may have changed in the meanwhile following the OTIP specifications. The password is never sent along the network as the proxy is running in the same host of the client process, thus it cannot be captured by an attacker possibly tracing the network traffic.

On the server side either an OTIP specific server program or a proxy interfacing an existing server daemon, provide the connectivity for OTIP clients. The Berkeley sockets and msocket API (application programming interface) use the `accept` system call to manage each TCP incoming connection `accept` takes as its first argument a socket descriptor used for listening to the arrival of new connections (listening socket) and returns a new socket descriptor connected to the remote client (connected socket) when a new connection arrives. When the address validity expires, OTIP closes the listening socket so that no new connections can take place using the old address, and it opens a new listening socket using the new address. Connected sockets continue to use the address which was the current one at their connection time. An address is completely dismissed when the last connection using that address gets closed.

It is easier to code an OTIP enabled server or an OTIP proxy using msocket API, as it is possible to start and close

 ISBN: 978-1-61208-305-6

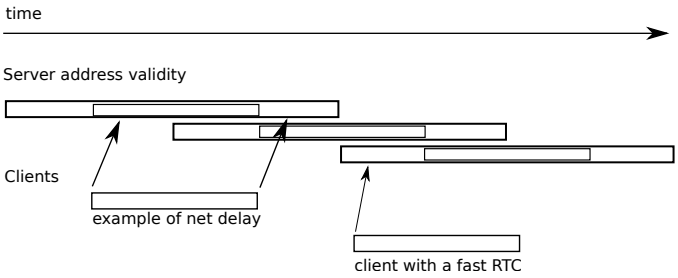


Fig. 3. Server address intervals of validity (as perceived from the server and by the clients)

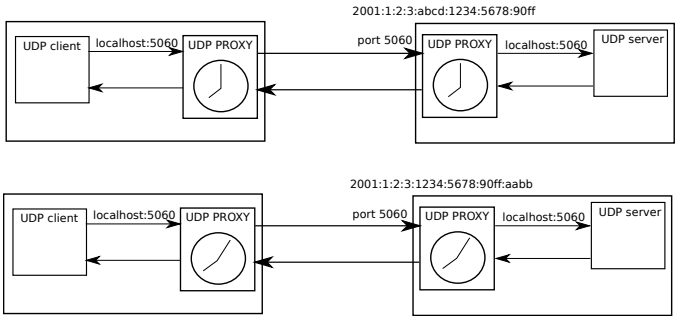


Fig. 4. A proposal to support OTIP for UDP services using existing clients and servers.

an entire TCP-IP stack each time a new address is needed or an old address expires. In the proof-of-concept implementation each address has a validity period of 64 seconds. On the server side each address is valid for 128 seconds. It is activated 32 seconds before the beginning of its validity period defined for the client and it expires 32 seconds after the end of the period as shown in Figure 3.

The overlapping between the validity periods of successive addresses has been set to 32 seconds, which should normally be a value far beyond the sum of the maximum difference between the clock readings plus the maximum transmission delay using NTP and a non-overloaded network.

In theory, this choice may reduce the time needed for a brute force attack by a factor less than four (two addresses always active for double the time). De-facto, 2^{62} attempts (which is $\sim 4\ 10^{18}$) to find an address which needs to be exploited in less than two minutes is a specification able to satisfy the security requirements of a wide range of applications.

The method described above for TCP cannot be directly applied to UDP. UDP is connection-less so it is not possible to keep the old address working only for the data exchanged along the existing connections. It is possible to write OTIP aware UDP clients which take care of the OTIP address validity periods and compute the new server address as needed. As a backwards compatibility feature, it is possible to implement OTIP UDP proxies which set up an OTIP UDP tunnel between the client and the server. Each client uses a port on the local host to exchange networking packets to the server. The local port is managed by the local proxy, which forwards each packet to the OTIP dynamically changing address of the proxy on the server side. On the server side, the OTIP UDP proxy forwards the packets to the OTIP unaware server using a local port (see Figure 4). Proxies are able to support several UDP clients at the same time. To do so, a new port is used by the client side proxies for each local client and by the server side proxies for each remote client. In this way the return packet can be properly rerouted.

VI. Limits, Future Developments and Conclusions

OTIP reduces the vulnerability of private services provided through the Internet. It makes it hard for attackers to discover the current address used by servers to communicate with their legitimate users. This method should be applied, together with other already known precautions, to protect communications along the Internet. A non-exhaustive list of such defenses, which should be put in place, includes the encryption of the communication payloads and the measures to avoid TCP spoofing, as defined in RFC 4953 [13]. In fact, an attack on the TCP sequence numbers could permit an attacker to *hijack* an existing TCP connection, avoiding, in this way, the OTIP protection.

Another scenario in which the protection of OTIP can be disrupted happens when even one user's host running the client program and the DNS proxy is compromised. An intruder who installed a root-kit on a host could read the OTIP passwords and compute the current addresses of the servers at a later time.

The implementation section above has shown how existing networking clients and servers can already use OTIP. Some programs may not be able to take advantage of this support. During the studies and tests we have seen two cases of programs which need some changes to use OTIP. In one case a TCP client did not properly managed the value of TTL returned by the DNS query and used a cached address far beyond its expiration. Unfortunately, such behavior is not very rare, as the server address is perceived by programmers as a constant. A second case happens when it was not possible to configure the client to change the address of the server for UDP communications, in order to use the UDP proxy. Sometimes the change of the server address is not a configurable entity in itself as it is the same address used for other services, or the UDP server address was communicated to the client as an element of the protocol.

As far as the address collision problem is concerned, we have shown its very limited and temporary impact. This drawback, however, cannot be easily eliminated by adding a duplicate address detection check, as proposed for the IPV6 privacy extensions [4], as servers and clients compute the current addresses independently. The clients cannot locally detect duplicate addresses on the servers' networks, and any sub protocol designed to force the clients to change the address sequence would weaken the methods, as it could also be used by attackers.

Although OTIP can be applied as described in section V of this paper, the syntax and implementation of OTIP should be standardized to provide a general purpose support to this new

 ISBN: 978-1-61208-305-6

feature. OTIP does not need to change any existing protocols, and it is able to support many existing client programs. Some standardization effort is needed, for example, to share the same OTIP DNS proxy for all the OTIP servers, as different service providers could use a different syntax or a different function to compute the current address.

It is worth noting that OTIP does not exclude hash based address definition as introduced in [12]. System administrators use hash based addresses to configure their hosts and DNS servers in an easier and less error-prone mode. Hash addresses can be used as base addresses for OTIP, combining a simple deployment of the network at the servers' side and the safety of the dynamic evolution of addresses as provided by OTIP.

The source code to test the experiments presented in this paper can be downloaded from `svn://svn.code.sf.net/p/view-os/code/branches/otiptest` and has been released under the GNU General Public License (GPL) v. 2 or newer. The programs are intended as just a proof-of-concept to show the effectiveness of the ideas introduced here.

REFERENCES

[1] D. Mills, J. Martin, J. Burbank, and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification," RFC 5905 (Proposed Standard), Internet Engineering Task Force, Jun. 2010.

[2] D. M'Raihi, S. Machani, M. Pei, and J. Rydell, "TOTP: Time-Based One-Time Password Algorithm," RFC 6238 (Informational), Internet Engineering Task Force, May 2011. [Online]. Available: http://www.ietf.org/rfc/rfc6238.txt (Retrieved: June 15, 2013)

[3] T. Narten and R. Draves, "Rfc 3041: Privacy extensions for stateless address autoconfiguration in ipv6," IETF, Tech. Rep., 2001.

[4] T. Narten, R. Draves, and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6," RFC 4941 (Draft Standard), Internet Engineering Task Force, Sep. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4941.txt (Retrieved: June 15, 2013)

[5] M. Tortonesi and R. Davoli, "User untraceability in next-generation internet: a proposal," in *Proceeding of Communication and Computer Networks 2002 (CCN 2002)*, IASTED, Ed., November 2002, pp. 177 – 182.

[6] R. Davoli, "Internet of threads," in *Proc. of the The Eighth International Conference on Internet and Web Applications and Services, ICIW 2013.*, 2013, pp. 100–105.

[7] ——, "LWIPV6," http://wiki.virtualsquare.org/wiki/index.php/LWIPV6 (Retrieved: June 15, 2013), 2007.

[8] L. Gardenghi, M. Goldweber, and R. Davoli, "View-os: A new unifying approach against the global view assumption," in *Proceedings of the 8th international conference on Computational Science, Part I*, ser. ICCS '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 287–296.

[9] R. Davoli and M. Goldweber, "msocket: multiple stack support for the berkeley socket api," in *SAC '12: Proceedings of the 27th Annual ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2012, pp. 588–593.

[10] LXC team, "lxc linux containers," http://lxc.sourceforge.net/ (Retrieved: June 15, 2013).

[11] POSIX.1-2008, "The Open Group Base Specifications," Also published as IEEE Std 1003.1-2008, San Francisco, CA, Jul. 2008.

[12] R. Davoli, "Ipv6 hash-based addresses for simple network deployment," in *Proc. of the The Fifth International Conference on Advances in Future Internet, AFIN 2013, To appear*, 2013.

[13] J. Touch, "Defending TCP Against Spoofing Attacks," RFC 4953 (Informational), Internet Engineering Task Force, Jul. 2007.

 ISBN: 978-1-61208-305-6

A Privacy-Enhanced User-Centric Identity and Access Management Based on Notary

Hendri Nogueira, Rick Lopes de Souza and Ricardo Felipe Custódio
Laboratory of Computer Security
Federal University of Santa Catarina
Florianópolis-SC, Brazil
Email: {jimi, rick.lopes, custodio}@inf.ufsc.br

***Abstract*—Identity and Access Management (IAM) systems aim to control of users' attributes for authentication, authorization and accountability processes. Public Key Certificates (PKCs), like the X.509 standard, use asymmetric key pairs to support digital signatures, authentication processes and to increase the trust in the communication. Nevertheless, the PKC does not concern itself with the management of users' attributes and their privacy to be used as an IAM system. We present a privacy-enhanced identity and access management architecture, addressing the user's management of his attributes and the privacy. With the user-centric paradigm and through the use of Identity-Based Cryptography (IBC), the model architecture is composed by a user-centric public key infrastructure. The asymmetric key pair enables the user to determine the control and the anonymity of his own attributes and the Notarial Authority validates the attributes claimed by the user. Our model aims for total control for the user in authentication and authorization procedures. Users can decide which attributes they want to disclose and which identity to use (e.g., real identity, pseudonym, anonymity).**

***Keywords*—User-centric; Identity Management; Notary; Attributes; Privacy-Enhancing; IBC.**

I. INTRODUCTION

Many standards related to Authentication and Authorization (AA) processes demand user's registration in the Services Providers (SPs), storing their attributes in the SPs database, and consequently replicating the attributes without users' control. Others divide responsibilities by those which manages uses' attributes and authenticate them, called the Identities Providers (IdPs), and those which only provide services for authorized users. Independently of that, the users' attributes need to be managed in a safe way and can not be used for other purposes than what was determined. Additionally, the AA systems must concern about the users' privacy and provide secure mechanisms to protect the users' identity and their related attributes.

The use of asymmetric cryptography keys have advantages in binding a key pair with the subject's attributes. The X.509 PKC, for example, can be applied to automated identification, authentication, digital signatures, access control and authorization functions in a digital environment [1], [2]. However, PKC is not recommended to be used for authorization procedures and when the user's privacy is a necessity. Additionally, the management of the PKCs by an X.509 Public Key Infrastructure (PKI) is difficult and expensive, requiring a lot of effort for its management and maintenance and leaving doubts as to the cost-benefit as regards its functionality [3]–[5].

The amount of verification procedure of a certificate and the revocation mechanism might be a disadvantage in some environments and situations with limited resources [6]. If the PKI is composed by many certification authorities and generating a large certification path until the end user certificate, the PKC's verification may not always be performed quickly. If the PKC's revocation constantly happens before the end of its validity, it interferes in the issuance of the certificates' revocation states in real-time.

Since most attributes for access control, role, and permission do not have a long lifetime (i.e., more than a certificate valid period), it is not recommended to include these types of attributes into a PKC. Moreover, an end user certification authority may not be the responsible for the management of those user's attributes, what means that the user's attributes values could be questionable. In this case, X.509 Attribute Certificates (ACs) could be a solution [7]. To provide a better security, PKCs and ACs should work together, but two different infrastructures are necessary to manage each one, PKI and X.509 Privilege Management Infrastructure (PMI) respectively [8]. However, this inherits the same issues from PKI and would increase the complexity, the costs, the human and computational resources.

The management of users' attributes in a PKC and how they are accessed do not concern about the user's privacy. As a PKC can be used as an off-line token to AA procedures, it needs to provide a sufficient amount of attributes to support the user in different situation. To reduce the amount of data in a unique PKC, an alternative could be the use of many PKCs with different attributes, but it is costly for the user. Furthermore, when a user provides his PKC, the verification system reads all the information in the certificate, even though what is not necessary for that procedure. Other privacy issue is related to the user's identity, which one (or more) identification attribute is used to bind with the user public key. Every time that the user uses the same PKC, the identification attribute and the public key bind to the action realized, and this can be traceable.

a) Contribution: Beyond the problems we have stated above, we present the concept of user-centric PKI architecture for identity and access management to improve the management, the disclosure, the users' control on their attributes and their privacy. The model explores two PKI problems: (1) the high costs of a PKC for end-users and leaving doubts as to the cost-benefit as regards for identity and access management, and (2) the X.509 PKC privacy deficiency, allowing the real iden-

 ISBN: 978-1-61208-305-6

tification of users, forcing users to reveal more attributes than needed, and enabling users' on-line transactions linkable across different websites. With this intention, our model introduce a new way that a user gets and uses an asymmetric cryptography key pair to claim his attributes to service providers and been validated by the Notary. Though the use of the identity-based cryptography and the user-centric paradigm, the user issues and manages their own private keys and issues self-signed assertions.

b) Outline: We start this paper by describing the related works. We also describe about privacy in IAM systems and we introduce about identity-based cryptography (with each correlated works), in Section III and Section IV, respectively. Next, we present our proposal followed by definitions (Section V). More practical descriptions of our idea, including a description of the procedures that players in our scheme perform are also shown (Section V-B). Afterwards, we describe some analysis about our model (Section VI). Finally, we present our considerations and future works (Section VII).

II. Related Work

Facing of the X.509 PKI problems described in the Introduction section, several works treat or propose alternatives to the revocation's problems. The Hormann et al.'s work, for example, aims at improving the existing revocation mechanism [9], while Scheibelhofer proposes a PKI without revocation checking and reducing the verification processes [10]. Faced with various revocation mechanisms, both existing and proposed, Ofigsbo et al. analyzed the cost of some mechanisms [11].

Alternative PKI models and concepts were created to give a different architecture of a PKI also. Focusing on digital signature issues, Moecke et al. [12] proposed a change to the form in which certificates are issued. Some optimizations were proposed, as in Vigil et al.'s [13] work. Vigil et al. [14] also proposed a new approach to X.509 PKI, based on notarials' responsibilities to support long-term signatures on documents. On the other hand, those works do not focus on users' attributes management and privacy.

The use of the notary responsibility is not new. Adams and Zuccherato's work [15] described a general notary service and protocols that notarial authorities validate signatures and provide up-to-date information regarding the status of certificates. This is also usable to extend the lifetime of a signature beyond key expiry or revocation. Another work based on notary is the Chao-yang's work that improved some computer notary system protocols to decrease the replay attack vulnerability in the agreement communication [16].

Another PKI scheme, the Simple Public Key Infrastructure (SPKI) [17] proposes and simplifies the PKI architecture and focuses on authorization processes, binding one key with a user's authorization [18]. Additionally, the Simple Distributed Security Infrastructure (SDSI) combines the SPKI design and the definition of groups to issue certificates to group membership [19]. SPKI/SDSI is limited because there is no formal bondage of trust between entities involved and a member can make an inquiry on behalf of its group.

III. Privacy in Identity and Access Management Systems

Concerning the existence of different specifications and frameworks for IAM systems, Jøsang and Pope's work [20] reports differences between the paradigms available. They concluded that the user-centric paradigm improves the user experience and the security of on-line service provision as a whole. Moreover, the user-centric paradigm aims the user's control at the different aspects of his identity, which it is used in different contexts and situations (called "partial identities"), and enhancing his privacy.

In on-line systems, where identities providers create access tokens on demand (e.g., SAML [21], OpenID [22], WS-Federation [23]) and also supporting a Single Sign-On (SSO) mechanism [24], they can lead to the impersonation of their users and the tracking of users' actions on-line. Systems with off-line token creation, such as X.509 certificates and some WS-Trust profiles [25] force the user to reveal more attributes than needed (as otherwise the issuer's signature cannot be verified) and make the on-line transactions linkable across different websites.

The privacy is made of terminologies, e.g., pseudonymity, anonymity, linkability, detectability, observability, and they provide different levels of privacy [26]. To point out two terminology above, anonymity and accountability are the extremes points related to the user linkability. Pseudonymity comprises all subset between and including the extremes above and all degrees of linkability. In each specific case, the strength of anonymity depends on the knowledge of certain parties about the linking relative to the chosen attacker model.

Privacy-enhancing identity management systems make the flow of personal data transparent and give users the control of their individual digital identity, i.e., their individual partial identities in an on-line world. The European PRIME project (Privacy and Identity Management for Europe) allows users to control the disclosure of their personal information and allows users to authenticate with anonymous credentials [27]. The PRIME architecture requires service providers to change their infrastructure server and the user needs to install the client side. The PRIME project succeeded to the PrimeLife (Privacy and Identity Management for Europe for Life) project. PrimeLife implemented the PrimeLife Policy Language [28].

Other anonymous credential system are the Idemix (short for "identity mix") [29] and the U-Prove [30]. The Idemix enables authentication, privacy and guarantees "anonymity" on the Internet. Nevertheless, the Idemix architecture is such complex and costly to implement for the issuer. The U-Prove specification uses cryptographic mechanisms which trusted parties issue "tokens" to users that contains user's attributes. The user is enable to select which attributes he wants to disclose from his "token" and the authentication could be in a anonymous way. However, U-Prove specification provides a revocation mechanism for the users' credentials by blacklisting the token identifier in which this turns the tokens linkable.

IV. Identity-Based Cryptography

Proposed by Shamir, the Identity-Based Cryptography (IBC) concept is based on the use of a string as a public key

 ISBN: 978-1-61208-305-6

for encryption and signature procedures [31]. The string is the user identity information (e.g., an email, name, IP). As a result, IBC significantly reduces the system complexity and the cost for establishing and managing the public key in a PKI [32], [33]. In 2001, Boneh and Franklin [34] and Cocks [35] solved the Shamir's identity-based encryption open problem. Baek et al. surveyed the state of research on identity-based cryptography [36].

Another approach is the Attribute-Based Signatures (ABS) that allows a party to sign a message with fine-grained control over identifying information. ABS is based on identity-based encryption in which each identity is considered as a set of descriptive attributes [37]. There are works use attribute-based signatures and attribute-based encryption to develop a cryptosystem for fine-grained sharing of encrypted data [38] and to propose a threshold attribute-based signatures (t-ABS) [39]. In our proposal, we use the concepts of IBC to compose our architecture's model and archive our goal.

In an IBC scheme, a trusted third party called Private Key Generator (PKG), aka Key Generation Center, is a trust authority responsible for generating the user's secret key. To be able to issue secret keys, the PKG needs to create a master secret key (*msk*) and the correspondent master public key (the public parameters and the public key itself) – *mpk*. The PKG's *mpk* is widely distributed and any party can compute a public key corresponding to an identity (*id*) by combining the master public key with the identity value. To get the corresponding secret key, it is necessary to authenticate through the PKG with that *id*. Then the PKG uses its master secret key and the user's *id* value to issue the corresponding secret key.

Some of the IBC advantages related to a standard PKI are: the public keys are derived from identifiers and than eliminates the need for a public key distribution infrastructure; the authenticity of the public keys is guaranteed implicitly as long as the transport of the secret keys to the corresponding user is kept secure; a compromised end-user secret key only exposes messages encrypted/signed with that particular *id* used to compute the secret key; no CRLs are needed; it is certificateless.

On the other hand, IBC also has disadvantages. Some of them are: a PKG needs to maintain a authentication infrastructure; the private key extraction has a very high exposure to man-of-the-middle attack; the PKGs do not interact with each other; it is necessary to support revocation of *ids* and consequently a well-defined expiry date for secret keys; and there is inherent key escrow, i.e., the users' secret key is known to the PKG.

V. User-Centric Public Key Infrastructure Based on Notaries

The User-Centric Public Key Infrastructure based on notaries (UCPKI) focuses on the management of users' attributes, where the user has more control and privacy over the disclose of his attributes to the services providers. UCPKI also addresses privacy-enhancement to the management of identity and access architecture, enabling anonymity, unlinkability and making the user untraceable. Based on the real world of notary responsibilities and services, the model's architecture has Notarial Authorities (NAs) that are trusted third parties responsible for verifying users' attributes as well as validating them. The model's architecture adopts the concept of identity-based cryptography and the user-centric paradigm in which the users issue and manage their own secret keys.

Considering that our model is user-centric and the IBC architecture needs a trust authority to issues secret keys based on identifications thus, it is the user who is going to realize that role, i.e., the role of a private key generator. As a consequence, the user maintains control of his identities used in each communication.

A. Components

In this subsection, we define the concepts involved in our model. We define two main entities: Attribute Registration Authority (ARA) and Notarial Authority (NA). UCPKI uses a Trust-service Status List (TSL) to keep the management of the trusted ARAs and to know the relation of each NA with the ARAs. Support to enhance the users' privacy is given by IBC.

1) Attribute Registration Authority: An Attribute Registration Authority is an entity responsible for registering attributes for the user (e.g., name, surname, e-mail address, occupation), storing the information in its trusted database system, and keeping attributes up to date. An ARA has to be responsible for, at least, one attribute from the user. Each ARA has an asymmetric cryptographic key pair to be used in the communication's workflow. The ARA's information and its public key are managed by a Trust-service Status List. Some examples of an ARA are the entities responsible for registering users' attributes for governmental, professional, or even business purposes.

2) Notarial Authority: A Notarial Authority is a point of trust responsible for receiving self-signed assertions from users and validating users' attributes. The NA communicates with the attribute registration authorities to confirm the correctness of the user's attributes. The validation of the assertion results in the assertion's signature by the NA (a co-signature). This procedure certifies the truthfulness of the user's attributes. To be defined as a trust authority, each NA has an asymmetric cryptographic key pair used to sign the assertions and to make the communication secure. The trust of the public keys tied to each NA and ARA is managed by a Trust-service Status List.

3) Trust-service Status List: A Trust-service Status List (TSL) is used to manage and inform the trust between NAs and ARAs. TSL turns trustworthy information about the entities relationships, along with a historical status and the associated public keys [40]. A TSL may be composed of a list of TSLs and it is managed, signed, and published into a public trust repository by a trusted entity of its domain.

B. How it Works

First, the user needs to create a master secret key and the correspondent master public key. To keep the *msk* safe, it is created in a secure device (e.g., smartcard or USB token) and it is protected with a PIN code. After the master key pair is created, the user must register his *mpk* in each ARA's database that manages at least one attribute about him. If the ARA already has an authentication mechanism installed,

 ISBN: 978-1-61208-305-6

then the registration of the user's *mpk* can be done after the user authentication. Otherwise, the most secure way is for the registration to be done personally.

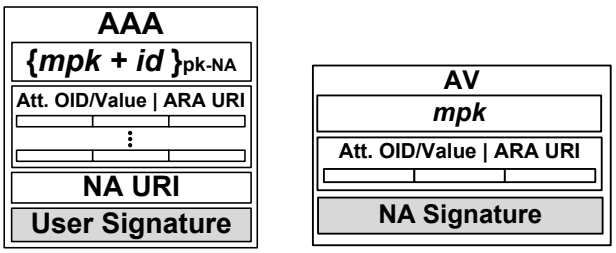
The validity of the master key pair is equally associated with the cryptographic algorithm used. If something were to happen to the user's *msk* during the time of validity, a procedure to change the registration of the user's *mpk* in the ARAs must be executed. For this change of *mpk* association in an ARA, we propose the use of a One Time Password (OTP) code [41] to facilitate the ARA's infrastructure and the user's life. In this case, the ARA does not necessarily have to maintain other authentication mechanism for the user (e.g., login and password), neither does the user need to remember his login information. The OTP code must be used only once and is given to the user after his *mpk* registration.

C. Accessing a Service Provider

To access an SP and get its resource, the user needs to choose an identity (e.g., real name, e-mail address, any string) and inform the necessaries attributes. The information is passed through a data structure, called the Attribute Authentication Assertion (AAA); see Figure 1a. Within an AAA, the user includes his (*mpk*) and his identifier ciphered with the public key of an NA ($\{mpk + id\}pk$-NA). This NA is chosen by the user preference. An AAA also contains: a set of attributes' Object Identifiers (OIDs), the attributes' values, and the referenced ARA responsible to the attributes (ARA URI); and the NA's reference (NA URI) to indicate which NA can correctly decipher the user's (*mpk* and identifier). The structure is signed by the user with the secret key corresponded to his chosen *id*.

Next, the user sends the AAA to the SP (illustrated in Figure 2 by step 1). The SP receives it and sends it (and also its public key) to the NA referenced in the AAA (step 2). The NA deciphers the user's *mpk* and identifier with its private key and uses the *mpk* with the user's *id* to verify the AAA's signature. If the signature is correct, the NA communicates to the referenced ARA to get the attributes verified (step 3). The NA sends the ARA a data structure, called Attribute Validation (AV) – see Figure 1b. An AV contains the user's *mpk* and the correspondent set of attributes' OIDs and values. Because it may have many attributes' sets related to the different ARAs, each set is verified through the correspondent *ARA URI*. All the communication is done by a secure channel to prevent the man-in-the-middle attack.

Each ARA manages the uses' attributes and the attributes are associated with the users' *mpk*. Therefore, when the ARA



(a) Attribute Authentication Assertion.

(b) Attribute Validation.

Fig. 1. Data structures used in the workflow model.

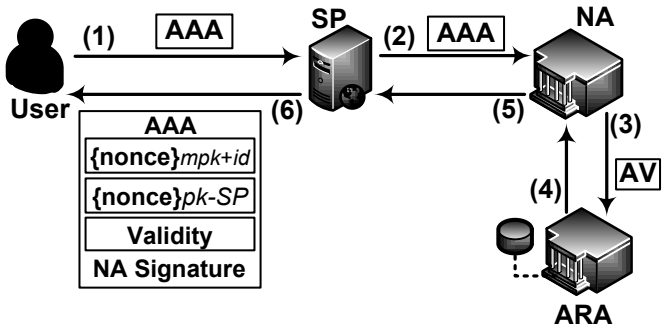


Fig. 2. Workflow to access a service provider.

receives NA's AV request, the ARA verifies the AV's signature and checks in its records if the associations of attributes' values are correct. If the ARA confirms the truth of the attributes, the ARA co-signs the AV and returns the signature as a confirmation response to the NA (step 4). After receiving all signatures from the ARAs involved, the NA generates a *nonce* to provide a challenge-response mechanism and the anonymous authentication of the user to the service provider. This *nonce* is ciphered with the user's *mpk* and the user's *id*. The NA also gets the same clear-text *nonce* and ciphers it with the SP's public key. Both nonces ciphered are attached to the AAA and then the AAA is co-signed by the NA with its private key. A validity period (e.g., a day, a week, a month) is also determined by the NA to indicate for how long those information are valid.

The co-signed AAA is sent back to the SP (step 5). The SP keeps a copy and the delivers the co-signed AAA to the user (step 6). Now, the user must authenticate (in a anonymous way) with the SP. This procedure is done by the use of the *nonce* created by the NA and included into the co-signed AAA. The user deciphers the *nonce* using the secret key related to the *id* used in the AAA. With the *nonce* in clear-text, the user ciphers again using the SP's public key and sends to the SP. The SP deciphers this cipher-text and gets the *nonce*'s value. The SP also deciphers the *nonce* included in the user's co-signed AAA and compares the two resulted values. If they were equal, the SP concludes that: the user who created the AAA is the same who has the master secret key (i.e., is the same user who created the secret key to sign the AAA with the related *id*); the attributes' values are validated through the NA; and the user is able to get the resources according to the SP's policies.

Once an AAA is co-signed by an NA, the user can reuse it with the same SP until the validity time included in the AAA. The AAA's validity could be based on the validity information included in the AAA or depending on the SP's policies. The SPs' key pair is managed by themselves and the public key is published publicly. Each NAs and ARAs' private key is managed in a secure device and the correspondent public key is managed in the TSL's domain.

VI. Analysis

The use of identity-based cryptography is essential to provide the dynamism and the facility to users in controlling which identities they want to use in each access. The IBC

procedures in our model eliminate the problems caused by the use of a public key certificate (cited in Section I) and also give the users more privacy to an identity and access management architecture. The key escrow provided by a common IBC is eliminated by the user-centric paradigm in our UCPKI model, in which the user maintains the total control of the master secret key and all secret keys related to each *id*.

The user's master secret key must be included into a secure device (e.g., token, smartcard) which the *msk* can not be moved, copied, and its usage must be protected by a password mechanism (e.g., PIN, OTP). The device should be able to realize cryptographic functions into it, like the generation of a secret key from an *id* and the signature of an AAA data structure. If the user loose his smartcard, he must do the procedure to change the registration of his *mpk* (as soon as possible) with all ARAs that manage his attributes.

The UCPKI architecture and the use of encryptions and signature procedures by the IBC increase the users' privacy through the secrecy of the users' identities, better management of their attributes, and the authenticity and integrity of the information's flows. The notarial authority contributes to increasing the security of the ARAs by limiting the ARAs' communication, which only the NAs would be able to request to verify the users' attributes. The NA also provides the users' attributes unlinkability, i.e., the SP can not link the user's identity inside the AAA with his attributes each time or each different services he accesses with different AAA (if the user determines a different *id* for each AAA). The ARA can not trace the user by analyzing each time the SP requests the user's attributes verification. The TSL manages the trust of the existed NAs and ARAs, keeping up to date their information and their public keys.

Anonymity and other privacy characteristics are also satisfied by the notarial authority, which is a trust entity and their policies must keep the security of the user's information during the procedures. The anonymity authentication procedure, through a *nonce* created and ciphered by the NA, provides the authenticity of the AAA sent by the user and the acknowledgement of the SP to confirm that the AAA was created by the same user with whom it is communicating. The AAA's signature done by the user (at the moment when the AAA is created) provides the authenticity, the integrity, and the non-repudiation, about the user's attributes claimed by himself. The signature made by the NA, co-signing the AAA, results in the veracity confirmation of the information claimed by the user, and that the attributes are binded to the cyphered user's master public key.

The user might store some AAAs already co-signed by the NA to speed up the process of requesting a resource to SP. With a co-signed AAA, the user could access a resource in an off-line mode, i.e., physically in the real world. To facilitate the AAAs' management, we assume that an application should be used to store the co-signed AAAs in a mobile device (with a secure mechanism) and the users' master secret key stored in a token and plugged into the device only when requested.

As a consequence of the ciphered *nonce* that is exchanged between the user and a service provider, each co-signed AAA works for a specific SP due to the *nonce* ciphered with the SP's public key. Another consequence of the proposed model is the transition of the responsibility's control of the attribute disclosed to their owners. It is important that the users being aware of how they should protect themselves when communicating with a service provider.

Differently from the traditional, already known, identity and access management systems, e.g., OpenID and SAML-based (like the Shibboleth framework [42]), the principal technology used in our model is the asymmetric cryptographic functions and it could also work in a non-web environment. Additionally, we do not propose a specific standard to be used in the communication's workflow neither we specify which technology must be used to implement the system. We only determine the paradigm, the concepts, the necessaries cryptographic functions, and letting the developer to decide which technology best fit for his implementation.

The differences between the UCPKI, Idemix and U-Prove user-centric approaches, mainly differ at the architecture. In the Idemix and U-Prove architectures, each attribute provider should be a credential issuer and there will be necessary a user authentication mechanism (e.g., login and password) to request the credential. The UCPKI one is based on notary, which it is responsible to communicate with the correspondent attribute provider to validate the user's attributes. Idemix and U-Prove are selective disclosure approaches, which many user's attributes are included into a smartcard and then, the user decides which ones will be disclosed at each use. At the UCPKI approach, each assertion has only those attributes that are going to be disclosed to that specific service provider. This approach provides a freshness of the user's attributes because the assertion does not need to have a long term validity.

VII. Considerations and Future Work

The use of the standard X.509 PKCs allows multiple digital processes becoming more secure for entities and information involved. However, this mechanism does not take into account the management of the users' attributes and their privacy. We presented a model that increases the way that users control and disclose their personal attributes. The UCPKI architecture aims to eliminates the complexity and problems caused by the PKI and PMI standards. The users' privacy is enhanced by the use of identity-based cryptography and the user-centric paradigm.

Based on the notaries' responsibilities, the notarial authorities validate the users' attributes communicating with the responsible attribute registration authority. The NAs increase the workflow and the users' privacy. Differently from other identity and access management infrastructures, UCPKI keeps the strength of the cryptography's functions and the dynamism of the IBC to simplify the authentication and authorization infrastructure. Additionally, UCPKI is less costly to end-users compared to PKI. For future works, we suggest a calculation of the processing necessities and the capabilities to focus in ubiquitous computing and environments. Moreover, the UCPKI model could be also applied in documents signatures procedures, and a description of the notarial authority validation procedures of the user's attributes and signature is needed to be compared with the PKCs ones.

 ISBN: 978-1-61208-305-6

REFERENCES

[1] D. Cooper *et al.*, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile," RFC 5280, Internet Engineering Task Force, May 2008.

[2] W. Diffie and M. Hellman, "New directions in cryptography," *Information Theory, IEEE Transactions on*, vol. 22, no. 6, pp. 644–654, Nov. 1976.

[3] P. Gutmann, "PKI: It's Not Dead, Just Resting," *Computer*, vol. 35, no. 8, pp. 41–49, Aug. 2002.

[4] A. Lioy, M. Marian, N. Moltchanova, and M. Pala, "PKI Past, Present and Future," *International Journal of Information Security*, vol. 5, pp. 18–29, 2006.

[5] C. Adams and M. Just, "PKI: Ten Years Later," in *In 3rd Annual PKI R&D Workshop*, 2004, pp. 69–84.

[6] D. Berbecaru, A. Lioy, and M. Marian, "On the Complexity of Public-Key Certificate Validation," in *Information Security*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2001, vol. 2200, pp. 183–203.

[7] S. Farrell, R. Housley, and S. Turner, "An Internet Attribute Certificate Profile for Authorization," RFC 5755, Internet Engineering Task Force, Jan. 2010.

[8] ITU-T, "Information Technology – Open Systems Interconnection – The Directory: Public-key and Attribute Certificate Frameworks," International Telecommunication Union - ITU, Tech. Rep., Nov. 2008, International Standard ISO/IEC 9594-8.

[9] T. P. Hormann, K. Wrona, and S. Holtmanns, "Evaluation of Certificate Validation Mechanisms," *Computer Communications*, vol. 29, no. 3, pp. 291–305, 2006.

[10] K. Scheibelhofer, "PKI without Revocation Checking," in *4th Annual PKI R&D Workshop*, NIST, Ed., 2005, pp. 48–61.

[11] M. Ofigsbo, S. Mjolsnes, P. Heegaard, and L. Nilsen, "Reducing the Cost of Certificate Revocation: A Case Study," in *Public Key Infrastructures, Services and Applications*. Springer Berlin Heidelberg, 2010, vol. 6391, pp. 51–66.

[12] C. T. Moecke, R. F. Custódio, J. G. Kohler, and M. C. Carlos, "Uma ICP Baseada em Certificados Digitais Autoassinados," in *SBSeg*, Fortaleza-CE, Brazil, 2010, pp. 91–104.

[13] M. A. G. Vigil, R. F. Custódio, N. da Silva, and R. Moraes, "Infraestrutura de Chaves Públicas Otimizada: Uma ICP de Suporte a Assinaturas Eficientes para Documentos Eletrônicos," in *SBSeg*, Campinas-SP, Brazil, 2009, pp. 129–142.

[14] M. A. G. Vigil, C. T. Moecke, R. F. Custódio, and M. Volkamer, "The Notary Based PKI – A Lightweight PKI for Long-term Signatures on Documents," in *EuroPKI*, Sep. 2012.

[15] C. Adams and R. Zuccherato, "Notary protocols," Internet Draft, Tech. Rep., 1997. [retrieved: Oct., 2013]. Available: http://tools.ietf.org/html/draft-adams-notary-01

[16] Z. Chao-yang, "An improved computer notary system protocols," in *Intelligence Information Processing and Trusted Computing (IPTC), 2011 2nd International Symposium on*, 2011, pp. 242–244.

[17] C. Ellison *et al.*, "SPKI Certificate Theory," RFC 2693, Internet Engineering Task Force, Sep. 1999.

[18] T. Saito, K. Umesawa, and H. Okuno, "Privacy enhanced access control by SPKI," in *Parallel and Distributed Systems: Workshops, Seventh International Conference on*, Oct. 2000, pp. 301–306.

[19] R. L. Rivest and B. Lampson, "SDSI – A Simple Distributed Security Infrastructure," Apr. 1996.

[20] A. Jøsang and S. Pope, "User Centric Identity Management," in *In Australian Computer Emergency Response Team Conference*, 2005.

[21] OASIS, "Conformance Requirements for the OASIS Security Assertion Markup Language (SAML) V2.0," OASIS Standard, Oct. 2005. [retrieved: Oct., 2013]. Available: http://docs.oasis-open.org/security/saml/v2.0/saml-conformance-2.0-os.pdf

[22] OpenID, "OpenID Authentication 2.0 - Final," Dec. 2007. [retrieved: Oct., 2013]. Available: http://openid.net/specs/openid-authentication-2_0.html

[23] OASIS, "Web Services Federation Language (WS-Federation) Version 1.2," OASIS Standard, 2009. [retrieved: Oct., 2013]. Available: http://docs.oasis-open.org/wsfed/federation/v1.2/os/ws-federation-1.2-spec-os.pdf

[24] H. Nogueira, D. B. Santos, and R. F. Custódio, "Um Survey sobre Ferramentas para Single Sign-On," in *Workshop de Gestão de Identidades - WGID/SBSeg*. Brazil: WGID/SBSeg, 2012, pp. 522–542.

[25] OASIS, "Oasis WS-Trust 1.4," OASIS Standard, Apr. 2012. [retrieved: Oct., 2013]. Available: http://docs.oasis-open.org/ws-sx/ws-trust/v1.4/ws-trust.pdf

[26] A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity,Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management," Aug. 2010, v0.34. [retrieved: Oct., 2013]. Available: http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf

[27] R. Leenes, J. Schallaböck, and M. Hansen, "Prime white paper," *PRIME (Privacy and Identity Management for Europe), White Paper*, 2008. [retrieved: Oct., 2013]. Available: https://www.prime-project.eu/prime_products/whitepaper/PRIME-Whitepaper-V3.pdf

[28] J. Angulo, S. Fischer-Hübner, E. Wästlund, and T. Pulls, "Towards usable privacy policy display & management for primelife," *Inf. Manag. Comput. Security*, vol. 20, pp. 4–17, 2012.

[29] J. Camenisch and A. Lysyanskaya, "An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation," in *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques: Advances in Cryptology*, ser. EUROCRYPT. Springer-Verlag, 2001, pp. 93–118.

[30] C. Paquin, "U-prove technology overview v1.1," Microsoft Corporation, Tech. Rep., April 2013. [retrieved: Oct., 2013]. Available: http://research.microsoft.com/pubs/166980/U-Prove%20Technology%20Overview%20V1.1%20Revision%202.pdf

[31] M. Joye and G. Neven, *Identity-Based Cryptography*. Ios PressInc, 2009, vol. 2.

[32] J. Oltsik, "The True Costs of E-mail Encryption," Enterprise Strategy Group, White Paper, 2010. [retrieved: Oct., 2013]. Available: http://www.trendmicro.de/media/ds/email-encryption-costs-esg-whitepaper-en.pdf

[33] A. Kumar and H. Lee, "Performance Comparison of Identity Based Encryption and Identity Based Signature," *International Journal of Security and Its Applications*, vol. 6, no. 3, pp. 19–28, 2012.

[34] D. Boneh and M. Franklin, "Identity-Based Encryption from the Weil Pairing," in *Advances in Cryptology – CRYPTO 2001*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2001, vol. 2139, pp. 213–229.

[35] C. Cocks, "An Identity Based Encryption Scheme Based on Quadratic Residues," in *Proceedings of the 8th IMA International Conference on Cryptography and Coding*. Springer-Verlag, 2001, pp. 360–363.

[36] J. Baek, J. Newmarch, R. Safavi-naini, and W. Susilo, "A Survey of Identity-Based Cryptography," in *Proc. of Australian Unix Users Group Annual Conference*, 2004, pp. 95–102.

[37] A. Sahai and B. Waters, "Fuzzy Identity-Based Encryption," in *Advances in Cryptology – EUROCRYPT 2005*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, vol. 3494, pp. 457–473.

[38] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based Encryption for Fine-grained Access Control of Encrypted Data," in *Proceedings of the 13th ACM conference on Computer and communications security*, ser. CCS '06. ACM, 2006, pp. 89–98.

[39] S. Shahandashti and R. Safavi-Naini, "Threshold Attribute-Based Signatures and Their Application to Anonymous Credential Systems," in *Progress in Cryptology – AFRICACRYPT 2009*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, vol. 5580, pp. 198–216.

[40] ETSI, "Electronic Signatures and Infrastructures (ESI); Provision of Harmonized Trust-service Status Information," Tech. Rep. TS 102 231, Dec. 2009.

[41] N. Haller, C. Metz, P. Nesser, and M. Straw, "A One-Time Password System," RFC 2289, Internet Engineering Task Force, Feb. 1998.

[42] H. Nogueira, R. F. Custódio, C. T. Moecke, and M. S. Wangham, "Using Notary Based Public Key Infrastructure in Shibboleth Federation," in *Workshop de Gestão de Identidades - WGID/SBSeg*. Brazil: SBSeg, 2011, pp. 405–414.

 ISBN: 978-1-61208-305-6

Resilient Delay Sensitive Load Management in Environment Crisis Messaging Systems

Ran Tao, Stefan Poslad, John Bigham
Dept. of Electronic Engineering and Computer Science
Queen Mary, University of London
London, UK
{ran.tao, stefan.poslad, john.bigham}@eecs.qmul.ac.uk

***Abstract*—Typical environment crisis messaging systems, e.g., those used in Tsunami Early Warning Systems, are open, distributed, and heterogeneous. In such systems, Publish Subscribe Message Oriented Middleware (PSMOM) is widely deployed using message brokers to enable open and distributed data publishers and subscribers to exchange raw and processed sensor data, authority driven workflows, and information generated by citizens. A key security challenge is that such message brokers may suffer a Denial of Service (DoS) attack, becoming overloaded and resulting in performance degradation or even worse in a broker crash. This significantly decreases the effectiveness of the system as vital messages may face unexpected delays or become lost. In order to address this challenge, a resilient workload management framework is required to better redistribute the message exchange from overloaded brokers to brokers with lesser loads. However, existing workload management mechanisms are not suitable to manage load in such environment crisis messaging systems as they are not designed to handle message traffic that may have different Quality of Service (QoS) requirements, e.g., different end-to-end transmission latency requirements. These may cause unexpected delays for sensitive messages or trigger unnecessary load balancing. In this paper, we propose a resilient delay sensitive workload management framework that extends an existing state-of-the-art messaging system, Publish/Subscribe Efficient Event Routing (PEER), by adding support for workload allocation, a Queue Depth load metric, and dynamic load thresholds, enabling end-to-end latency guarantees and avoiding unnecessary load balancing. The model has been validated in a simulation.**

Keywords-PSMOM; Denial of Service attack; Workload Management

I. INTRODUCTION

Modern environment crisis management systems, such as Tsunami Early Warning Systems (EWS) follow a System-of-System (SoS) framework that integrates various messaging components and subsystems, e.g., different information sources, processing services, and crisis simulation systems, and takes into account the open, distributed, heterogeneous, and collaborative nature of such systems. In such a SoS framework, PSMOM is deployed as a messaging bus because it allows components and subsystems to be distributed on heterogeneous platforms and to communicate asynchronously in a loosely coupled manner [1]. In addition, QoS-aware policies can be used to help differentiate message traffic in a PSMOM to allow different types of data, such as raw and processed sensor data, service data, and simulation data to be exchanged via inter-linked message brokers [13]. Figure 1 shows an example EWS framework based on PSMOM (Messaging Bus) support. In this framework, "P" are message publishers or pubs that label messages with respect to different subjects and send these to message Brokers "B". "S" are subscribers or subs that request the messages of interest to them and receive messages matched to their interests via a message broker. The message interaction in the system consists of the following. First, a sensor data bus type broker acquires physical sensor data from different sources, e.g., physical sensors, such as buoys and tide gauges, and human sensed data via social networks data sources, such as Twitter on mobile phones. Second, a database (DB) receives and records the live sensor data and publishes the historical sensor data via a processing message broker or bus. Third, this processing bus receives both live and historical sensor data, processes this data and publishes the analysis results to a User Interface (UI). Fourth, the UI receives and displays the analysis results. Fifth, a service controller publishes service control messages to message components when they need to change its performance to adapt to a changing environment situation, e.g., to increase the sensor data collection frequency in case the onset of a crisis is detected. With the support of PSMOM, these system components can be distributed in monitor centres at different geographically locations and work collaboratively.

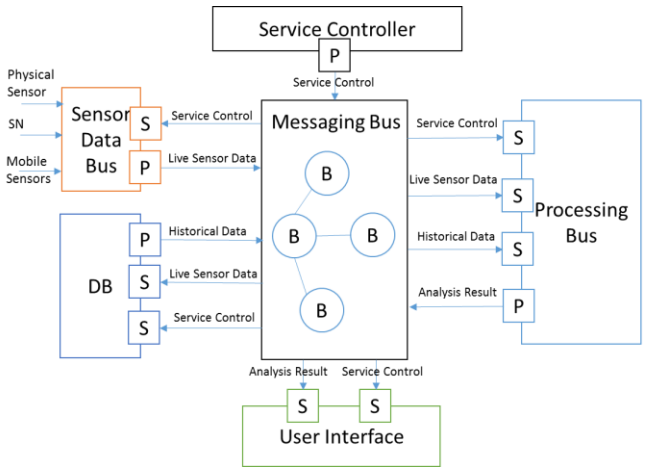


Figure 1. EWS with PSMOM Support

 ISBN: 978-1-61208-305-6

A core security risk in such environment crisis messaging systems is a Denial of Service attack [13] that significantly reduces the efficiency and accuracy of an early warning when message brokers become overloaded. This can be caused by: rogue publishers that can flood the broker with large fake messages, high-rate messages, and many useless topics; rogue subscribers with a slow subscription speed can cause messages to build up in the broker. A standard method to avoid such problem is to use user authorization, i.e., only authorized pubs and subs are legal and able to exchange messages using the message broker. However, this blocks the unauthorized publishers that could come online and provide useful information to improve the ground truth at a crisis. The above attacks can be modeled as a message burst (rogue publisher attack) and a capacity reduction (rogue subscriber attack). Workload management through an improved broker resilience model, e.g., mirroring and load balancing, is a feasible solution. Some forms of resilience, such as mirroring, are quite standard and are already supported in our resilient messaging system. Instead, in this paper, we focus on a more challenging workload management sub-system to provide load balancing for message brokers in EWS.

Existing MOM workload management mechanisms are not applicable in EWS because of the following limitations. First, much work focuses on homogeneous broker models where brokers are assumed to have the same processing power and bandwidth. However, EWSs tend to be heterogeneous because different system components and subsystems have varying CPU, memory, disk size and network bandwidth. Second, the heterogeneity of messages is not fully considered. Although messages have been divided into different subjects and assigned with different sizes and rates, different QoS requirements for different types of messages are ignored. This may trigger unnecessary load balancing and result in a waste of system resources or introduce unexpected delays to time-critical messages and result in a delay for critical decision-making.

In this paper, we propose a delay sensitive workload management solution for PSMOM used in EWS. This solution extends the Publish/Subscribe Efficient Event Routing (PEER) framework [1] by adding a workload distribution mechanism that assigns message brokers with least utilized load capacities to clients, a Queue Depth load metric and dynamic thresholds, to provide latency guarantees and to avoid unnecessary load balancing.

The remainder of the paper is organized as follows. Section II describes related work. Section III shows the system overview. Section IV describes the workload management framework. Section V presents a validation of the framework. Section VI reports the conclusions and projects the future work.

II. RELATED WORK

Load balancing in distributed system has been widely researched for over two decades [1, 4]. The goal of load-balancing solutions is to efficiently distribute the workload to the available resources so as to lower the risk of system overload and to maintain system performance.

Load balancing solutions can be executed in different layers: the network layer, operating system layer, middleware layer, and application layer. The layer, where the load balancing mechanisms can effectively detect and balance the load, is the best place to deploy the solution. For example, it would be ineffective to use a random DNS redirection strategy in the network layer or perform process migration in the OS layer for load balancing. This is because these approaches cannot identify the relationship between subscriptions nor estimate the load imposed by a subscription onto a broker [1]. Therefore, we focus on the load balancing strategy in the middleware layer as a PSMOM system is middleware based.

In a PSMOM system, the broker workload depends on the number and type of subscriptions served by this broker, i.e., on message size and the incoming and outgoing messages rates. Load balancing in a PSMOM is achieved by migrating subscriptions from overloaded brokers to ones with lesser loads.

Gupta et al. [6] proposed two types of load balancing in a peer-to-peer content-based PS system [2, 12]. Load balancing is achieved by splitting the peer with the heaviest subscription load in half and propagating events to a newly joint replicated peer. Chen and Schwan [7] proposed an optimized overlay reconstruction algorithm that performs load distribution based on CPU load. Load Balancing is triggered only when clients find a broker that is closer than its current connected broker. Subscription clustering [8, 9, 10, 11] is another solution that partitions a set of subscriptions into a number of clusters in order to reduce the overall network traffic. The above solutions can balance the load but they are all designed for homogeneous systems.

Cheung et al. [1] proposed the PEER framework that aims to overcome the above limitations for load balancing in PSMOM. Its primary target is content-based PSMOM but the author claims that it can also be applied to topic-based PSMOM. In PEER, brokers have different processing capabilities and Internet links. The load of a broker is detected by periodically monitoring three middleware layer load metrics: input utilization, matching delay, and output utilization, and comparing the monitoring results of each metric with two static thresholds. Among these metrics, input utilization is determined by the quotient of the input rate (R_{input}) in messages per second over the matching rate ($R_{matching}$) in messages per second, i.e., $R_{input}/R_{matching}$; matching delay is defined as the average time (in second) spent in a broker to process matching; output utilization is defined as the quotient of the used bandwidth (BW_{used}) over the total bandwidth (BW_{total}), i.e., BW_{used}/BW_{total}. If unbalanced load or overload is detected, a load balancing is triggered and the system migrate subscriptions from the offloading broker onto a load-accepting broker, while not overloading it. An evaluation of the design compared to a naive random load balancing approach shows that PEER is capable of efficiently balancing load in a heterogeneous messaging environment. However, PEER ignores the heterogeneity of system applications, such as the different end-to-end latency requirements, and therefore may trigger unnecessary load balancing if all the applications are delay

 ISBN: 978-1-61208-305-6

tolerant or introduce unexpected delays for delay sensitive applications. In addition, it does not distinguish the uplink that is used to disseminate messages to subscribers and downlink that is used to receive messages. The differences between these may introduce different client migration priorities in a load-balancing phase. Further, there is no pre-emptive workload distribution mechanism in PEER. It therefore requires extra work to migrate subscriber clients from one (edge) broker to another based on the load differences.

Our work extends the PEER framework by adding support for workload allocation and more comprehensive delay sensitive aware load detection, and redesigns the load analysis and balancing mechanisms to fit the detection and distribution mechanism.

III. SYSTEM OVERVIEW

In Figure 1, multiple message brokers (B) form a messaging bus that works as an integrated message exchange. In our design, these brokers are organized into a Head-Edge Broker model that is motivated by the architecture adopted by Google's distributed publish/subscribe system GooPS for use in MOM deployments in real world applications [1]. Our design targets enhancements to the Head-Edge Broker model (Section III.A) by providing delay sensitive load management supported with management agents (Section III.B).

A. Head-Edge Broker Model

The Head-Edge broker model (H-E model) organizes the brokers into a hierarchy structure, as shown in Figure 2.

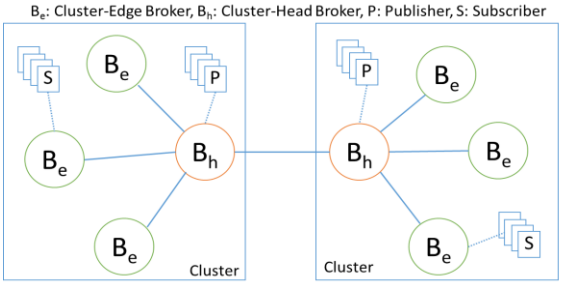


Figure 2. PEER Head-Edge Broker Model

A broker with more than one neighbour broker is referred to as a cluster-head broker (B_h), while a broker with only one neighbour broker is referred to as a cluster-edge broker (B_e). A cluster-head broker together with its connected edge brokers form a cluster. In the H-E model, publishers are only served by B_h, and subscribers are only served by B_e, so that, in a cluster, messages are always routed from the B_h to B_e. Inter-cluster message dissemination is achieved by having a B_h forwarding publication messages to the B_h of all matching clusters.

B. Management Agent

To manage the workload for H-E model, a Management Agent (MA) is allocated for each broker. The MA belonging to the head broker is called HMA, while the one belonging to the edge broker is named EMA. Both HMA and EMA consist of an Overlay Manager (HOM and EOM respectively), a Load Detector (HLD and ELD), and a Load Analyser (HLA and ELA).

HOM receives the broker allocation request from all the clients in the cluster and assigns brokers to the clients according to the client's source (a publisher or a subscriber), the availability of the broker, and the distribution status of existing clients. In addition, when load balancing is triggered, HOM notifies selected clients to migrate from original brokers to the new load-accepting brokers. What's more, HOM interacts with EMA to update the load information of edge brokers, and interacts with HOM of other clusters to share the cluster-based load information. ***EOM*** updates the load status of the edge broker to HOM and receives the load status of other edge brokers in the same cluster from HOM. In addition, when load balancing is triggered, EOM updates the available selected subscriptions to the HOM. Both HOM and EOM work with its relevant load analysers to generate an offloading client list that contains the clients to be migrated from the overloaded broker to the load-accepting broker when load balancing is required.

HLD and ***ELD*** detect the load status, e.g., as a set of fuzzy states, LOW, HIGH, and OVERLOAD, of the relevant broker, i.e., HLD monitors the head broker and ELD monitors edge brokers. Although the authors in [1] claim that the head broker is less likely to be overloaded since it does no matching work for subscribers, the head broker can become overloaded when it reaches its maximum network capacity whilst exchanging messages. So, HLD monitors the network bandwidth used by the head broker and reports its status to HOM when its load state changes. ELD does similar work but it needs to monitor all the load metrics (see section IV.B) and report this to the EOM. In addition, to get the dynamic threshold, ELD periodically detects the transmission latency between the edge broker and head broker, between subscribers and edge brokers, and request HLD to detect the transmission latency between publishers and head brokers.

HLA and ***ELA*** analyse the load distribution for clients, e.g., the Internet usage of individual client, store the observations into a table and pass this to the relevant OM. In addition, the clients in the overloaded broker are prioritized for offloading when its load metric exceeds its threshold otherwise making the broker become overloaded.

IV. LOAD MANAGEMENT FRAMEWORK

In this design, the workload management framework consists of a workload distribution phase, a load detection phase, and a load-balancing phase. In the workload distribution phase, HOM allocates brokers to each new subscriber based on the load status of the edge brokers and the distribution of existing subscribers. In the load detection phase, the load of the broker is periodically detected and the change of the load status is updated and sent to its OM. During the load balancing phase, a three step offloading strategy is adopted, i.e., locating the load-accepting broker(s), selecting subscriptions, and migrating the selected

 ISBN: 978-1-61208-305-6

subscriptions from the overloaded broker to the load-accepting broker(s).

A. Workload Distribution

In practice, it is very important to avoid the OVERLOAD problem by optimizing the workload distribution beforehand. In this workload management framework, the workload distribution process is designed using the following principles:

- First, subscribers of the same topic are allocated to the same broker to avoid extra network bandwidth usage, as same messages are no longer routed to different edge brokers.
- Second, topics that are highly correlated are allocated to different brokers [5], as they may introduce a sudden increase in broker load.
- Third, new subscriber clients are allocated to brokers that have the least utilized load capacity that is computed from all the load metrics (Section IV.B).

B. Load Detection

To accurately detect the load status of a broker, the load metric and related thresholds need to be clarified. In the H-E broker model, brokers are classified into a cluster-head broker and cluster-edge broker, and different load metrics and thresholds are allocated to the different types of brokers.

1) Load Metrics for Head Broker and Edge Broker

The main tasks of a B_h are: to route messages from publishers and B_h of other clusters to the B_e that serves matched subscribers; to route messages from publishers to the B_h of another clusters that serve matched subscribers. As claimed in [1], a B_h is less likely to be overloaded for doing the matching work as no subscribers connect to it. Therefore, the load status of a B_h is mainly affected by the network bandwidth usage. Table I lists the load metrics used for cluster-head broker.

TABLE I. LOAD METRICS FOR THE HEAD BROKER

Metric	Expression
Downlink Utilization	Input-Rate / Downlink-Bandwidth
Uplink Utilization	Output-Rate / Uplink-Bandwidth

B_e serves all subscribers, and therefore does a lot more matching work. So, the load matching costs need to be monitored. In addition, since a guaranteed end-to-end transmission delay is required, a Queue Depth metric that measures the number of messages waiting in the output queue and reflects the message waiting time in a broker is introduced. Table II lists the load metrics used for cluster edge broker.

TABLE II. LOAD METRICS FOR EDGE BROKER

Metric	Expression
Downlink Utilization	Input-Rate / Downlink-Bandwidth
Matching Utilization	Input-Rate / Matching-Rate
Uplink Utilization	Output-Rate / Uplink-Bandwidth
Queue Depth	No. of Messages waiting in each Output Queue

2) Threshold Determination

We introduce two thresholds for each metric to describe the load status of a broker. A lower threshold (TH_{low}) indicates whether or not a broker is available to accept more loads, while a higher threshold (TH_{high}) indicates whether or not load shifting is required. Based on the two thresholds, the load status of a broker is divided into LOW LOAD, HIGH LOAD, and OVERLOAD. The relationship between the threshold and the load status is defined in Table III.

TABLE III. LOAD STATE & THRESHOLD

Condition	Status
(All the metrics) $< TH_{low}$	LOW LOAD
$TH_{low} <$ *(Any metric) & (All the metrics)* $< TH_{high}$	HIGH LOAD
$TH_{high} <$ *(Any metric)*	OVERLOAD

The higher value the HIGH LOAD threshold is set to (e.g., 99% CPU Utilization), the more the system resources can be used. However, a broker can become overloaded before it can do any offloading. The magnitude of the difference between the lower and higher threshold controls the efficiency of load balancing and the level of the load imbalance between brokers. For example, a small difference, e.g., 1%, reduces the load imbalance between brokers but makes brokers more likely to enter OVERLOAD from HIGH LOAD, which may result in endless load balancing cycles [1]. In addition, based on whether or not the load metrics are affected by the delay sensitivity of the messages, the load metrics are divided into two groups and assigned with different thresholds.

Both uplink usage and downlink usage for B_h are set with static thresholds, i.e., TH_{low} = 0.9 and TH_{high} = 0.95. The same thresholds are applied to the downlink utilization and the matching utilization for the edge broker. These values are retrieved from the threshold defined for PEER [1]. The uplink usage and the Queue Depth metric of a B_e are considered separately as they affect the time of messages waiting in the broker. In this design, only TH_{low} is assigned to the uplink utilization metric of the edge broker as it is only used to indicate whether or not the broker is available for more loads, and only TH_{high} is set for the Queue Depth metric that is used to trigger load balancing with latency guarantees. The value of TH_{low} for the uplink utilization of the edge broker is set the same as others, e.g., 0.9, while the value of TH_{high} for Queue Depth of edge broker is calculated based on the end-to-end latency requirements for different topics of individual subscribers, the transmission delays, and the time a message spent in brokers. The following procedure shows the steps of determining the dynamic TH_{high} for Queue Depth metric.

a) Transmission Time

The end-to-end latency requirement for subscriber "s" on topic "T" is denoted as $t_{s,T}$. The practical end-to-end latency is calculated as the sum of the total transmission time (t_{s,T_trans}) and the total time spent in broker (t_{s,T_broker}). With the H-E model, the total transmission time is obtained based on the transmission time from publishers to B_h ($t_{s,T\_p\text{-}h}$), from B_h to B_h of matching clusters ($t_{s,T\_h\text{-}h}$), from B_h to B_e of the

matched subscribers ($t_{s,T\_h\text{-}e}$), and from B_e to subscribers ($t_{s,T\_h\text{-}s}$), i.e., $t_{s,T\_trans} = t_{s,T\_p\text{-}h} + t_{s,T\_h\text{-}h} + t_{s,T\_h\text{-}e} + t_{s,T\_e\text{-}s}$. For the case that publisher clients on the same topic are served by different clusters, the transmission time obtained for different publishers may have different values since the time cost from publishers to B_h and from B_h to B_h may be different. In our design, the maximum transmission time from all the obtained transmission time is selected, denoted as $t_{s,T\_trans\text{-}sel}$.

b) Time in Broker

The total time spent in broker (t_{s,T_broker}) consists of the time spent in B_h that serves the publisher (t_{s,T_h}), the time spent in the remote B_h belonging to the matched clusters ($t_{s,T\_remote\text{-}h}$), the B_e that serves the matched subscribers ($t_{s,T\text{-}e}$), i.e., $t_{s,T\_broker} = t_{s,T\_h} + t_{s,T\_remote\text{-}h} + t_{s,T\_e}$. For each broker, the time cost is the sum of the arrival time ($t_{s,T\_arrival}$), departure time ($t_{s,T\_departure}$), the matching time ($t_{s,T\_matching}$) and the time waiting in the queue ($t_{s,T\_waiting}$). Each of the arrival and departure time is determined by the size of the message and the uplink/downlink bandwidth, and the matching time is mainly affected by the number of filters in the matching process. The waiting time in a broker is determined by the number of messages waiting in the queue and the message output rate.

c) Dynamic Threshold

With the end-to-end transmission delay, the maximum time that a message can spend in the output queue of broker B_e ($t_{s,T\text{-}e}$) can be determined as $t_{s,T} - t_{s,T\_trans\text{-}sel} - t_{s,T\_h} - t_{s,T\_remote\text{-}h} - t_{s,T\_arrival\text{-}e} - t_{s,T\_matching\text{-}e} - t_{s,T\_departure\text{-}e}$. This maximum-allowed time a message can spend in the output queue varies due to the change of transmission time, matching time and arrival/departure time. This maximum waiting time in the message broker is used to compute the higher threshold for Queue Depth metric for subscriber "s" on topic "T", i.e., the value of Queue Depth at a time t_i ($QD_{s,T}(t_i)$) must follow the condition defined in (1), where $\lambda_{s,T}(t_{i+1})$ and $\mu_{s,T}(t_{i+1})$ are the predicted message input rate and output rate in message/s for time t_{i+1}, and t_{LB} is the average time cost for load balancing that is mainly affected by the notification message transmission time from HOM to subscribers, e.g., from milliseconds to seconds, and the analysis time, e.g., in milliseconds.

$$\frac{QD_{s,T}(t_i) + [\lambda_{s,T}(t_{i+1}) - \mu_{s,T}(t_{i+1})]}{\mu_{s,T}(t_{i+1})} \le t_{s,T-e} - t_{LB} \quad (1)$$

Therefore, the higher threshold for Queue Depth at time t_i ($TH_{s,T}(t_i)$) is found with (2).

$$TH_{s,T}(t_i) = \left(t_{s,T\text{-}e} - t_{LB}\right) * \mu_{s,T}(t_{i+1}) - [I_{s,T}(t_{i+1}) - \mu_{s,T}(t_{i+1})] \quad (2)$$

C. Load Analysis

Load analysis is invoked when a broker is overloaded. A load analyser aims to estimate and profile the load distribution for individual clients served by the broker, and prioritizes the offloading clients according to their overloaded load metrics.

1) Load Estimation

Both ELA and HLA compute the network bandwidth usage for individual clients based on the message exchange rate and the bandwidth, e.g., the uplink usage of edge broker for subscriber "s" on topic "T" is computed as the message output rate ($\mu_{s,T}$) / uplink bandwidth. In addition, ELA estimates the matching utilization and records the Queue Depth for each subscriber on each topic.

2) Priorities Offloading Client

In our design, the clients of the same topic are recognized as a bundle in the offloading process, i.e., they are either migrated together to the load-accepting broker or kept together in the overloaded broker. Only if the load-accepting broker cannot accept any bundle of clients, these clients are dealt with separately.

In the head broker, the publishers of different topics can be categorized into four groups: the publishers that only have remote subscribers (P_r), the publishers that have both local and remote subscribers ($P_{r\text{-}l}$), the publishers that only have local subscribers (P_l), and the publishers that have no subscribers (P_n). So, if the broker is in a downlink overload state, the priority of all the publishers are $P_n > P_r > P_{r\text{-}l} > P_l$, while if it is an uplink overload state, the priority relationship becomes $P_r > P_{r\text{-}l} > P_l > P_n$. The difference between the two is the location of P_n, because migrating publishers with no subscribers cannot reduce the uplink utilization but only reduce the downlink utilization.

In each edge broker, similar to the equivalent situation with head brokers, the subscribers on different topics can be categorized into S_r, $S_{r\text{-}l}$, S_l and S_n. In addition, for the Queue Depth metric, as it does not relate to the locations of the publishers, the subscribers are categorized into three groups: subscribers without message waiting in the queue (S_{empty}), subscribers with message waiting in the queue but not overloaded ($S_{w\text{-}no}$), and subscribers of which the Queue Depth metric is overloaded ($S_{overload}$). In all the groups above, the subscribers are ordered based on its allowed waiting time, i.e., the larger the waiting time, the higher the priority. The relationship between subscribers is defined in Table IV.

TABLE IV. PRORITIES SUBSCRIBERS IN EDGE BROKER

Overload Metric	Priority
Downlink Utilization	$S_r > S_{r\text{-}l} > S_l > S_n$.
Uplink Utilization	
Matching Utilization	$S_n > S_r = S_{r\text{-}l} = S_l$
Queue Depth	$S_{empty} > S_{w\text{-}no} > S_{overload}$

D. Load Balancing

After the load analysis process, load balancing takes place. As described in PEER, if a head broker becomes overloaded, load balancing happens between head brokers in different clusters by migrating publishers from an overloaded head broker to head brokers with lesser loads. If instead, the edge broker becomes overloaded, the load balancing first takes place within a local cluster. Only if there is no available load-accepting broker in the local cluster, i.e., no broker is in the LOW LOAD state, or the available load-accepting brokers have less load capacity than that required

 ISBN: 978-1-61208-305-6

by the overloaded broker to recover from OVERLOAD state, is inter-domain load balancing invoked. All the load balancing processes follow a similar three-step offloading strategy, i.e., Load-Accepting Broker Locating, Client Selection, and Client Migration. In this paper, intra-domain load balancing between edge brokers is described below as an example.

1) Load-Accepting Broker Locating

The EOM of an overloaded broker checks the load state of brokers in the same domain to locate brokers in a LOW LOAD state and sends a load balancing request to a HOM with the candidate broker ID(s). The HOM records whenever a broker is in a load-balancing phase and sends requests to all candidate brokers. The EOMs of these candidate brokers report the values of all the load metrics to the HOM. And when HOM receives this information, it will in turn forward to the requesting EOM.

2) Client Selection

Based on the results of step 1, EOM of the overloaded broker prioritizes the candidate brokers based on the value of the overloaded load metric of the broker, i.e., the broker with the lowest value of the load metric has the highest priority to accept the load. In addition, from the prioritized client list, EOM retrieves the clients and estimates the load influence to the load-accepting broker for all the load metrics, e.g., for the uplink bandwidth usage, the influence is estimated as the (input rate of the client / the uplink bandwidth of the load-accepting broker), which means that if the clients are migrated to the load-accepting broker, the uplink usage will be increased by this amount. So, in the case that the client does not overload the load-accepting broker, it is selected and put in an offloading list. The selection process continues until the estimated load status of the overload broker is not OVERLOAD any more. The offloading list is then sent to the HOM. HOM notifies the EOMs of the selected edge brokers to be in a load-balancing phase.

3) Client Migration

In the last step, HOM sends messages to all the clients that are in the offloading list, asking them to start a message exchange via the load-accepting broker(s). All the clients then set up connection(s) to the load-accepting broker(s) and drop the connection to the offloading broker except for subscribers that have messages waiting in the queue. In this case, the subscribers will drop the connections only when all the messages waiting in the overloaded broker are received. In addition, a message is sent by each client to HOM to confirm the completion of the migration process. HOM counts the number of clients that have completed the migration away from the overloaded broker. There is also a default timeout for the migration so that the load-balancing phase can stop even if some clients stop the message exchange during the migration. When all the clients complete the migration or the waiting time has timed out, the HOM notifies all the EOMs involved in the load-balancing phase that the load balancing is complete.

V. VALIDATION

We validate our framework by comparing our load balancing mechanism to that designed for the PEER framework. In this paper, a local load balancing triggered by Queue Depth metric is given as an example. The setup used for the local load balancing experiment involves four edge brokers (B_0, B_1, B_2, and B_3) connected to one cluster-head broker (B_h) to form a star topology, which forms a messaging bus to exchange information in an EWS. The simulation environment specification is listed in Table V. For each broker, the uplink bandwidth and downlink bandwidth is the same and is static during the experiment so that the broker-to-broker transmission latency will not change, e.g., is set at 0.1s. In addition, we assume that the client to broker transmission latency is also constant during the experiment, e.g., 0.2s.

TABLE V. SIMULATION EXPERIMENT SPECIFICATION

Broker ID	Specifications		
	CPU (MHz)	*Memory (MB)*	*Bandwidth (Mbps)*
B_h	2000	64	20
B_0	800	32	6.5
B_1	1500	32	8
B_2	1300	64	5
B_3	1000	64	8

Messages for 15 topics are published, i.e., in the EWS system, 15 types of data are exchanged through the messaging bus. The number of publishers for each topic is a random number, e.g., 1-5. Each publisher publishes messages in an average rate of 50 message/s. The number of subscribers for each topic is a random number, e.g., 1-8. In the experiment, we assume that subscribers of different topics have different end-to-end latency requirements but the subscribers of the same topic have the same requirement. The average message size changes for different topics, e.g., from 200 Byte to 1KB. Table VI gives an example of how topics are specified in one experiment.

TABLE VI. TOPIC SPECIFICATIONS IN ONE EXPERIMENT

Topic ID	No. of Pubs	No. of Subs	Latency Requirement (s)	Msg Size (Byte)
1	1	8	1.8	200
2	2	2	1.7	800
3	5	1	1.6	1000
4	4	2	1.5	400
5	3	3	1.4	200
6	1	1	1.3	400
7	2	5	1.2	300
8	2	7	1.1	400
9	5	2	1.0	500
10	4	4	0.9	200
11	1	5	30	600
12	3	2	60	400
13	1	6	40	200
14	2	3	50	300
15	4	5	100	200

The reason to use a random number is to allow the broker loads to be varied in different experiments to improve the validation. On the other hand, the reason to have such a

 ISBN: 978-1-61208-305-6

range, e.g., 1-5 for publisher, is to lower the chance that all the brokers become overloaded since in that case load balancing is not useful - more brokers are required.

According to the end-to-end transmission latency requirements and the assumptions for the static client-to-broker and broker-to-broker transmission delay, the maximum time of a message can be held in a broker can be determined, e.g., for topic 1, $t_{topic1\text{-}broker} = 1.8 - 0.2 - 0.1 = 1.5s$. These values are used in experiment to determine higher threshold for the Queue Depth metric.

In experiment start-up, all brokers are instantiated simultaneously with the MAs. After that, all publishers register and connect to head brokers, and MAs start to measure the load status of a broker and the broker-to-broker transmission delays. Each experiment is divided into three phases: 1) client distribution phase: 1s – 15s, subscribers of each topic in EWS are registered and distributed to the available brokers in each second; 2) equilibrium phase: 15s - 29s, both publishers and subscribers in EWS are running without message bursts and client joining or leaving; 3) message burst simulation and load balancing phase: at 30s, a burst that simulates a message flood when a crisis detected is generated by doubling the speed of publishing 7 topics (e.g., topic 2, 4, 5,..., 12, 14); after 31s, up to the end of the experiment, load balancing will be triggered if any load metric exceeds its higher threshold. The reason to set time slots to these values is to highlight the changes in each stage of the simulation. The experiment can be easily expanded by 1) adding more brokers, publishers and subscribers; 2) increasing the time intervals for each phase; 3) generating more message bursts.

Figure 3 shows the simulation results for the uplink utilization in percent (y) against time in second (x). The value above 100% indicates that the output queue starts to build up.

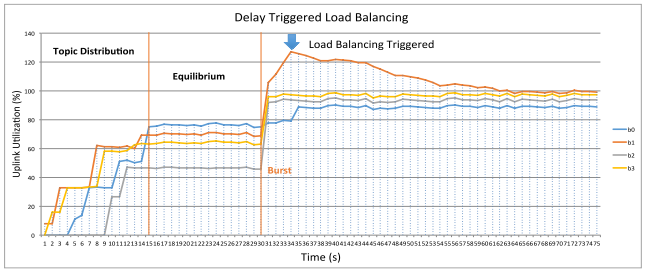


Figure 3. Simulation Result for Uplink Utilization

After the workload distribution, broker b1 serves topics 1, 3, 8, and 14 (refer to the 4 inflection points of b1 in the topic distribution stage). In addition, for b1, the output queue starts to build up after a message burst (30s) as the uplink utilization exceeds 100%; 4s after this (34s), the queue depth value of topic 8 exceeds the TH_{high}, and thus load balancing is triggered. Topic 1 in b1 is migrated to broker b0. Therefore, broker b1 has more bandwidth to clear the messages for topic 8 in the queue (from 34s – 62s, a balancing stage). After 62s, the message queue for topic 8 in broker b1 is removed. The uplink utilizations for all the brokers are below 100%. Figure 4 shows the Queue Depth, i.e., number of messages in the output queue, for topic 8 in broker b1.

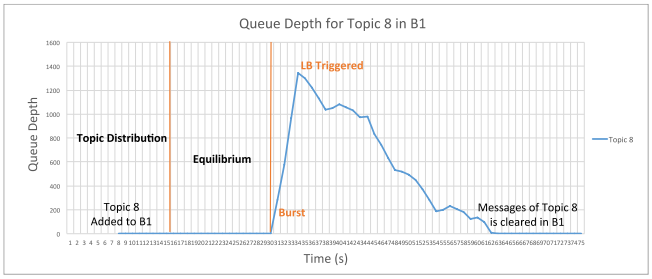


Figure 4. Queue Depth for Topic 8 in broker b1

When the same simulation is applied using PEER load balancing mechanism, the results are shown in Figure 5.

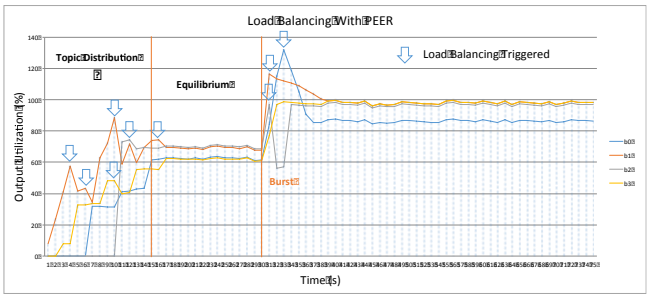


Figure 5. Simulation Result for Uplink Utilization for PEER

In the topic distribution phase (1s-15s) of PEER, all the subscribers are initially connected to broker b1 and migrated to other brokers (e.g., b2) based on the load differences (as there is no work distribution mechanism in their work). In addition, after a burst (30s), as the delay requirements for topics are ignored, unnecessary load balancing takes place between broker b0 and b2 (at time 31s), that results in an additional load balancing to balance the two at time 33s. Comparing Figure 3 to Figure 5, the differences indicate that our proposed delay-aware load balancing method is more effective in workload distribution, and can avoid unnecessary load balancing as the delay requirements are considered.

VI. CONCLUSION AND FUTURE WORK

In this paper, an analysis of existing load management solutions for PSMOM was presented. Existing solutions ignored the end-to-end delay requirements, which may introduce unexpected delays for delay sensitive messages or trigger unnecessary load balancing that introduces extra overhead to the system, and therefore they were not applicable for PSMOM in EWS. To address the above limitations, we proposed a delay sensitive load management solution that extends an existing state-of-the-art, PEER framework [1]. In addition, an intra-cluster load balancing example was presented with comparison to PEER and the results showed that the proposed framework is aware of the delay requirements, and has the potential to efficiently solve the broker overload problem in a LAN-based setting.

The framework was implemented with Apache Qpid [14], an open source AMQP based MOM product. In the

 ISBN: 978-1-61208-305-6

future, real sensor data from the TRIDEC project will be adopted to evaluate the framework in a WAN-based setting.

ACKNOWLEDGMENT

This work is supported in part by the EU FP7 funded project TRIDEC (FP7-258723-TRIDEC) and by a PhD studentship at Queen Mary University of London.

REFERENCES

[1] A. K. Y. Cheung and H.-A. Jacobsen, "Load Balancing Content-based Publish/Subscribe Systems", ACM Transactions on Computer Systems, Vol. 28, Issue 4, Article 9, 55 pages, Dec. 2010, doi: 10.1145/1880018.1880020.

[2] I. Aekaterinidis and P. Triantafillou, "PastryStrings: A Comprehensive Content-Based Publish/Subscribe DHT Network", 26th IEEE International Conference on Distributed Computing Systems (ICDCS 06), Jul. 2006, pp. 23-32, doi:10.1109/ICDCS.2006.63.

[3] P. Tran and P. Greenfield, "Behavior and Performance of Message-Oriented Middleware Systems", Proc. 22nd International Conference on Distributed Computing Systems Workshops (ICDCSW 02), Jul. 2002, pp. 645-650, doi:10.1109/ICDCSW.2002.1030842.

[4] A. K. Y. Cheung and H.-A. Jacobsen, "Dynamic Load Balancing in Distributed Content-based Publish/Subscribe", Proc. 7th ACM/IFIP/USENIX International Conference on Middleware (Middleware 06), Nov. 2006, pp. 141-161.

[5] J. Wang, J. Bigham, and J. Wu, "Enhance Resilience and QoS Awareness in Message Oriented Middleware for Mission Critical Applications", 8th International Conference on Information Technology: New Generations (ITNG 11), Apr. 2011, pp. 677-682, doi: 10.1109/ITNG.2011.120.

[6] A. Gupta, O. D. Sahin, D. Agrawal, and A. E. Abbadi, "Meghdoot: Content-based Publish/Subscribe over P2P Network", Proc. 5th ACM/IFIP/USEaNIX International Conference on Middleware (Middleware 04), Oct. 2004, pp. 254-273.

[7] Y. Chen and K. Schwan, "Opportunistic Overlays: Efficient Content Delivery in Mobile Ad Hoc Networks", Proc. 6th ACM/IFIP/USENIX International Conference on Middleware (Middleware 05), Nov. 2005, pp. 354-374, doi: 10.1007/11587552_18.

[8] E. Casalicchio and F. Morabito, "Distributed Subscriptions Clustering with Limited Knowledge Sharing for Content-Based Publish/Subscribe Systems", 6th IEEE International Symposium on Network Computing and Applications (NCA 07), Jul. 2007, pp. 105-112, doi: 10.1109/NCA.2007.16.

[9] A. Riabov, Z. Liu, J. L. Wolf, P. S. Yu, and L. Zhang, "Clustering Algorithms for Content-Based Publication-Subscription Systems", Proc. 22nd International Conference on Distributed Computing Systems (ICDCS 02), May 2002, pp. 133-142, doi:10.1109/ICDCS.2002.1022250.

[10] A. Riabov, Z. Liu, J. L. Wolf, P. S. Yu, and Li Zhang, "New Algorithms for Content-Based Publication-Subscription Systems", Proc. 23rd International Conference on Distributed Computing Systems (ICDCS 03), May 2003, pp. 678-686, doi: 10.1109/ICDCS.2003.1203519.

[11] T. Wong, R. H. Katz, and S. McCanne, "An Evaluation of Preference Clustering in Large-scale Multicast Applications", 9th IEEE International Conference on Computer Communications (INFOCOM 00), Mar. 2000, pp. 451-460, doi:10.1109/INFCOM.2000.832218.

[12] Y. Zhu, "Ferry: A P2P-Based Architecture for Content-Based Publish/Subscribe Services", IEEE Transactions on Parallel and Distributed Systems, Vol. 18, May 2007, pp. 672-685, doi:10.1109/TPDS.2007.1012.

[13] F. Paganelli, G. Vannuccini, D. Parlanti, D. Giuli, and P. Cianchi, "GEMOM Middleware Self-healing and Fault-tolerance: a Highway Tolling Case Study," The Sixth International Conference on Systems and Networks Communications (ICSNC 11), Oct. 2011, pp. 136-142.

[14] Apache Qpid, Official Web Page, http://qpid.apache.org (last access date: September 24th, 2013)

 ISBN: 978-1-61208-305-6