ICSEA 2014

The Ninth International Conference on Software Engineering Advances

ISBN: 978-1-61208-367-4

October 12 - 16, 2014

Nice, France

ICSEA 2014 Editors

Herwig Mannaert, University of Antwerp, Belgium

Luigi Lavazza, Università dell'Insubria - Varese, Italy

Roy Oberhauser, Aalen University, Germany

Mira Kajko-Mattsson, Stockholm University & Royal Institute of Technology, Sweden

Michael Gebhart, iteratec GmbH, Germany

ICSNC 2014

Forward

The Ninth International Conference on Systems and Networks Communications (ICSNC 2014), held between October 12 - 16, 2014 in Nice, France, continued a series of events covering a broad spectrum of systems and networks related topics.

The conference covered fundamentals on wireless, high speed, sensor and 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:

- Multimedia communications systems
- High speed networks
- Wireless networks
- Sensor and vehicular networks
- Security and P2P systems

Similar to the previous edition, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the ICSNC 2014 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to ICSNC 2014. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the ICSNC 2014 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope ICSNC 2014 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of systems and networks communications. We also hope that Nice, France provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

ICSNC 2014 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 2014 Research Institute Liaison Chairs

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

ICSNC 2014 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 2014 Special Area Chairs

Mobility / vehicular
Maode Ma, Nanyang Technology University, Singapore

Pervasive education
Maiga Chang, Athabasca University, Canada

ICSNC 2014

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 2014 Research Institute Liaison Chairs

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

ICSNC 2014 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 2014 Special Area Chairs

Mobility / vehicular
Maode Ma, Nanyang Technology University, Singapore

Pervasive education
Maiga Chang, Athabasca University, Canada

ICSNC 2014 Technical Program Committee

Habtamu Abie, Norwegian Computing Center - Oslo, Norway
Fakhrul Alam, Massey University, New Zealand
Jose M. Alcaraz Calero, University of the West of Scotland, UK
Pedro Alexandre S. Gonçalves, Escola Superior de Tecnologia e Gestão de Águeda, Lisbon
Abdul Alim, Imperial College London, UK
Shin'ichi Arakawa, Osaka University, Japan
Seon Yeob Baek, The Attached Institute of ETRI, Korea

Hong Shen, University of Adelaide, Australia
Roman Shtykh, CyberAgent, Inc., Japan
Axel Sikora, University of Applied Sciences Offenburg, Germany
Adão Silva, University of Aveiro / Institute of Telecommunications, Portugal
Narasimha K. Shashidhar, Sam Houston State University, USA
Stelios Sotiriadis, University of Derby, UK
Theodora Souliou, National Technical University of Athens, Greece
Tim Strayer, BBN Technologies, USA
Weilian Su, Naval Postgraduate School - Monterey, USA
Masashi Sugano, Osaka Prefecture University, Japan
Young-Joo Suh, Pohang University of Science and Technology (POSTECH), Korea
Jani Suomalainen, VTT Technical Research Centre of Finland, Finland
Stephanie Teufel, University of Fribourg, Switzerland
Radu Tomoiaga, University Politehnica of Timisoara, Romania
Neeta Trivedi, Neeta Trivedi, Aeronautical Development Establishment- Bangalore, India
Tzu-Chieh Tsai, National Chengchi University, Taiwan
Thrasyvoulos Tsiatsos, Aristotle University of Thessaloniki, Greece
Manos Varvarigos, University of Patras, Greece
Costas Vassilakis, University of Peloponnese, Greece
Luis Veiga, INESC ID / Technical University of Lisbon, Portugal
Tingkai Wang, London Metropolitan University, UK
Alexander Wijesinha, Towson University, USA
Riaan Wolhuter, Universiteit Stellenbosch University, South Africa
Ouri Wolfson, University of Illinois, USA
Mengjun Xie, University of Arkansas at Little Rock, USA
Erkan Yüksel, Istanbul University - Istanbul, Turkey
Yasir Zaki, New York University Abu Dhabi, United Arab Emirates
Sherali Zeadally, University of Kentucky, USA
Weihua Zhang, Fudan University, China

Copyright Information

Table of Contents

A Novel Spectrum Sensing Scheme for Dynamic Traffic Environments

Keunhong Chae and Seokho Yoon
College of Information and Communication Engineering, Sungkyunkwan University,
300 Cheoncheon–dong, Jangan–gu, Suwon–si, Gyeonggi–do, 440–746, Korea
Email: syoon@skku.edu

***Abstract*—This paper proposes a novel spectrum sensing scheme for dynamic traffic circumstances, where a Primary User (PU) signal may randomly depart or arrive during the sensing period, and thus, the conventional schemes developed under the assumption of the static traffic circumstances perform poorly. In the proposed scheme, a sensing period is partitioned into several sub-periods and the final decision on the existence of the PU signal is made by combining sensing decisions during each sub-period. From numerical results, it is confirmed that the proposed scheme has an improved detection performance compared with those of the conventional schemes.**

Keywords*—*Spectrum sensing; Cognitive radio; Dynamic traffic circumstance; Detection probability

I. Introduction

The spectrum sensing determines the existence of a Primary User (PU) signal on frequency bands and thus enables the Cognitive Radio (CR) to allocate a vacant frequency band to a Secondary User (SU) without interfering the PU [1][2]. So far, most spectrum sensing techniques [3]-[5] have been developed under the assumption of static traffic circumstances where the PU is present or absent during the whole sensing period [6]. In practical wireless communications, however, the PU traffic could be dynamic, i.e., the PU could arrive at or depart from the frequency band in the middle of a sensing period randomly, and in such a dynamic traffic circumstance, the conventional spectrum sensing techniques where the PU traffic is assumed to be static would perform poorly [5].

An interesting spectrum sensing scheme [7] was proposed for dynamic traffic circumstances, where the dynamic behavior of the PU signals is modeled as a Poisson random process and an improvement in detection performance over the conventional schemes was shown. However, the scheme of [7] requires additional information including the arrival and departure rates of PU signals and the performance improvement is not significant especially in low Signal-to-Noise Ratios (SNRs) of practical interest. In this paper, we propose a novel spectrum sensing scheme based on partitioning of a sensing period, where a sensing period is partitioned into several sub-periods and the final decision on the existence of a PU signal is made by combining sensing decisions during each sub-period. Numerical results demonstrate that the proposed scheme provides a better detection performance than those of the conventional schemes.

The rest of this paper is organized as follows. Section II introduces the received signal model of the CR systems in the dynamic traffic circumstances. In Section III, we explain the proposed spectrum sensing scheme based on sensing period partitioning. The proposed scheme is confirmed to perform better than the conventional ones in Section IV, and finally, Section V concludes the paper.

II. System model

In the dynamic traffic circumstances, the spectrum sensing problem can be modeled as a binary hypothesis testing problem with the null hypothesis H_0 and the alternative hypothesis H_1:

$$H_0 : y[n] = \begin{cases} x[n]+w[n] & \text{for } n = 1,\ 2,\ ...,\ J_0, \\ w[n] & \text{for } n = J_0+1,\ J_0+2,\ ...,\ N, \end{cases} \tag{1}$$

and

$$H_1 : y[n] = \begin{cases} w[n] & \text{for } n = 1,\ 2,\ ...,\ J_1, \\ x[n]+w[n] & \text{for } n = J_1+1,\ J_1+2,\ ...,\ N, \end{cases} \tag{2}$$

where $y[n]$ is the nth sample of the received signal, $x[n]$ is the nth sample of the PU signal, $w[n]$ is the nth additive white Gaussian noise sample, N is the number of observed samples, and the hypothesis H_0 (H_1) represents that a PU signal is present (absent) on the frequency band at the beginning of a sensing period and then the PU signal departs from (arrives at) the band between the J_0th (J_1th) and the (J_0+1)th ((J_1+1)th) samples. In this paper, we assume that the departure or arrival of PU signals follows a Poisson random process and occurs only once in the sensing period, as in [7].

III. Proposed spectrum sensing scheme

A. Test Statistics

To solve the hypothesis testing problem in the dynamic traffic circumstances, first, we partition a sensing period composed of N observed samples into K sub-periods with N/K samples each, where N is assumed to be divisible by K and a set of samples in the kth sub-period is denoted by

$$G_k = \{y[n]\}_{n=1+(k-1)N/K}^{k(N/K)}. \tag{3}$$

Then, we perform energy detection to determine the existence of a PU signal during each sub-period with a threshold λ_G predetermined by a given false alarm probability. Thus, the test statistic T_{G_k} corresponding to the kth sub-period is obtained as

$$T_{G_k} = \sum_{n=1+(k-1)N/K}^{k(N/K)} |y[n]|^2,\ k = 1,\ 2,\ 3,\ ...,\ K. \tag{4}$$

Comparing T_{G_k} with λ_G, finally, we decide the existence of the PU signal during the kth sub-period:

$$T_{G_k} \underset{D_{0,\,G_k}}{\overset{D_{1,\,G_k}}{\gtrless}} \lambda_G,\ \text{for}\, k = 1,\ 2,\ ...\ ,\ K, \tag{5}$$

where $D_{1,\,G_k}$ ($D_{0,\,G_k}$) represents that the PU is present (absent) during the kth sub-period.

If a PU signal arrives at (departs from) the frequency band at the end of a sensing period in a dynamic traffic circumstance, the probability that the whole energy in a sensing period

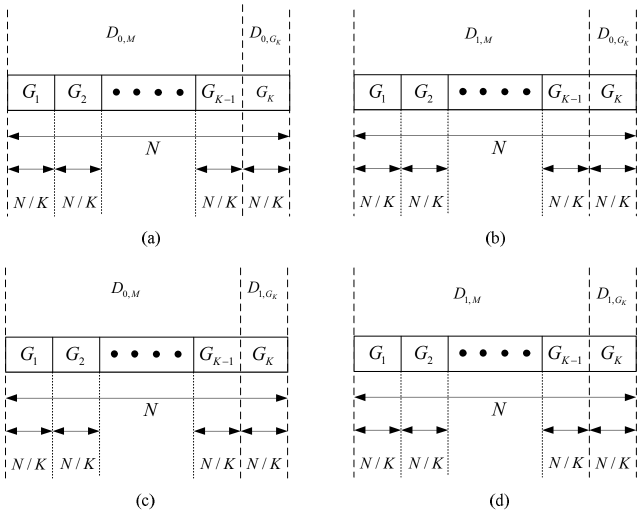


Fig. 1. Four cases of $(D_{x,\,M},\, D_{y,\,G_K})$.

exceeds the threshold would be small (large) even if the PU signal is actually present (absent) at the end of the sensing period, and thus, the conventional spectrum sensing techniques using simply the whole energy during a sensing period would yield a wrong decision on the existence of the PU signal. To solve this problem, we use the sensing decisions from K sub-periods as follows. Depending on the presence or absence of the PU signal in each sub-period, firstly, we represents the kth sensing decision x_{G_k} as 1 or 0, i.e.,

$$x_{G_k} = \begin{cases} 1 \text{ when } T_{G_k} > \lambda_G \\ 0 \text{ when } T_{G_k} \le \lambda_G \end{cases} \qquad (6)$$

for $k = 1, \cdots, K-1$. Using the majority rule, subsequently, we make an intermediate-decision as

$$\sum_{k=1}^{K-1} x_{G_k} \underset{D_{0,\,M}}{\overset{D_{1,M}}{\gtrless}} \frac{K-1}{2}, \qquad (7)$$

where the intermediate decision $D_{1,\,M}$ $(D_{0,\,M})$ represents that the PU is present (absent) during the first $K-1$ sub-periods and it is combined with the sensing decision $D_{1,\,G_K}$ or $D_{0,\,G_K}$ corresponding to the Kth sub-period, yielding the final decision on the existence of the PU signal. We have four different combinations $(D_{1,\,M},\, D_{0,\,G_K})$, $(D_{0,\,M},\, D_{0,\,G_K})$, $(D_{0,\,M},\, D_{1,\,G_K})$, and $(D_{1,\,M},\, D_{1,\,G_K})$ as shown in Figure 1 and it is easy to see that the hypotheses H_0 and H_1 are finally declared for $(D_{1,\,M},\, D_{0,\,G_K})$ and $(D_{0,\,M},\, D_{0,\,G_K})$ and for $(D_{0,\,M},\, D_{1,\,G_K})$, respectively. On the other hand, as shown in Figure 2, we have two cases corresponding to H_0 (Figure 2(a)) and H_1 (Figure 2(b)), respectively, for $(D_{1,\,M},\, D_{1,\,G_k})$. Thus, the final decision is made by comparing the average of the test statistics $\{T_{G_k}\}_{k=1}^{K-1}$ with the test statistic T_{G_K} of the Kth sub-period for those cases. Finally, the decision rule can be summarized as

$$\begin{cases} H_0 \text{ for } (D_{0,M},\, D_{0,G_K}) \text{ and } (D_{1,M},\, D_{0,G_K}) \\ H_1 \text{ for } (D_{0,M},\, D_{1,G_K}) \\ \frac{\sum_{k=1}^{K-1} T_{G_k}}{K-1} \underset{H_0}{\overset{H_1}{\lessgtr}} T_{G_K} \text{ for } (D_{1,M},\, D_{1,G_K}). \end{cases} \qquad (8)$$

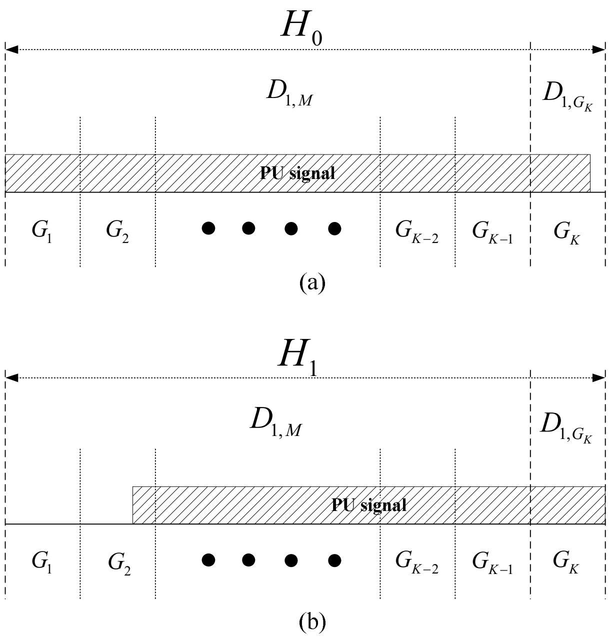

H_0
$D_{1,M}$
D_{1,G_K}
PU signal
G_1 G_2 • • • • G_{K-2} G_{K-1} G_K
(a)
H_1
$D_{1,M}$
D_{1,G_K}
PU signal
G_1 G_2 • • • • G_{K-2} G_{K-1} G_K
(b)

Fig. 2. Random departure and arrival cases of $(D_{1,\,M},\, D_{1,\,G_K})$.

In other words, in case of $(D_{1,\,M},\, D_{1,\,G_k})$, the proposed scheme compares the T_{G_K} with average of $\{T_{G_k}\}_{k=1}^{K-1}$. In other cases, the proposed scheme decides the existence of the PU signal based on the detection result of sub-period G_K.

B. *Distribution of Test Statistics*

In this section, we derive the probability density function (PDF) of the test statistic T_{G_k}. Assuming that the additive white Gaussian noise samples $\{w[n]\}_{n=1}^{N}$ have zero-mean and unit-variance, we can easily see that T_{G_k} follows the non-central chi-square distribution

$$p_{T_1}(T_{G_k}) = \left(\frac{T_{G_k}}{2s_k^2}\right)^{\frac{(\frac{N}{K}-2)}{4}} e^{\frac{-(s_k^2+T_{G_k})}{2}} I_{\frac{N}{2K}-1}(s_k\sqrt{T_{G_k}}) \qquad (9)$$

when a PU signal is present, where

$$s_k^2 = \sum_{n=1+(k-1)N/K}^{k(N/K)} x^2[n] \qquad (10)$$

and

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{\alpha+2k}}{k!\Gamma(\alpha+k+1)} \qquad (11)$$

is the αth-order modified Bessel function of the first kind with $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ and $x > 0$.

When a PU signal is absent, on the other hand, $y[n] = w[n]$ and thus T_{G_k} obeys the central chi-square distribution

$$p_{T_2}(T_{G_k}) = \frac{1}{2^{\frac{N}{2K}}\Gamma(\frac{N}{2K})} T_{G_k}^{\frac{N}{2K}-1} e^{-\frac{T_{G_k}}{2}}. \qquad (12)$$

IV. Numerical results

In this section, we compare the spectrum sensing performances of the proposed and conventional [3][7] schemes in terms of the detection probability defined as Pr$(H_1|H_1)$. For simulations, N is set to 200 and the SNR is defined as σ_s^2/σ_n^2,

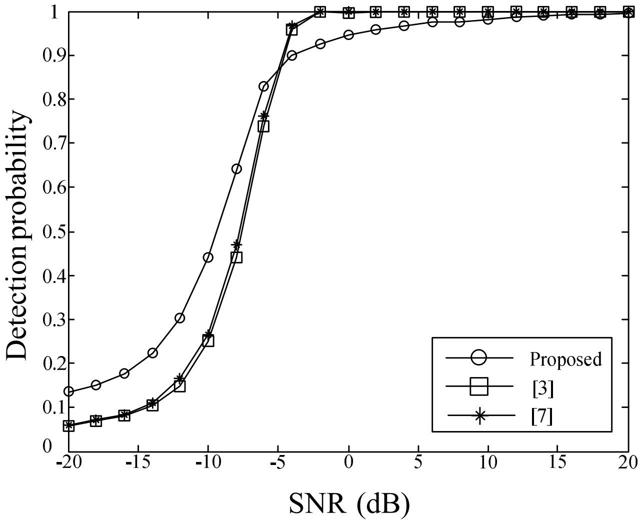


Fig. 3. The detection probability as a function of the Signal-to-Noise Ratio when $P_{fa} = 0.05$.

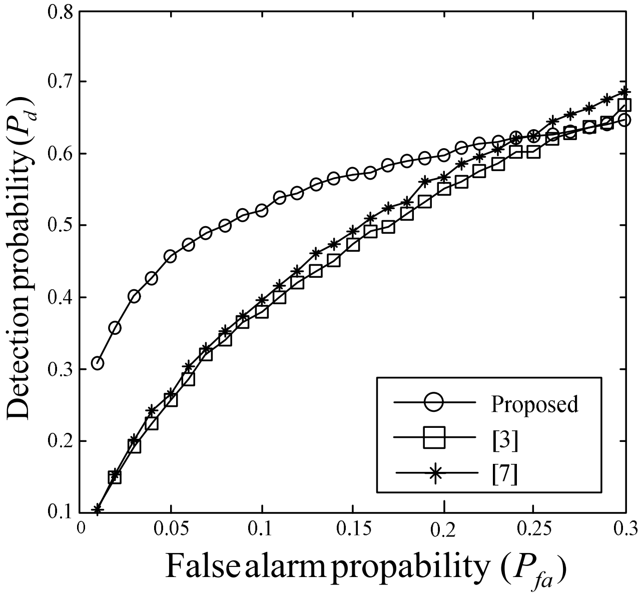


Fig. 4. The ROC curves when SNR= -10 dB.

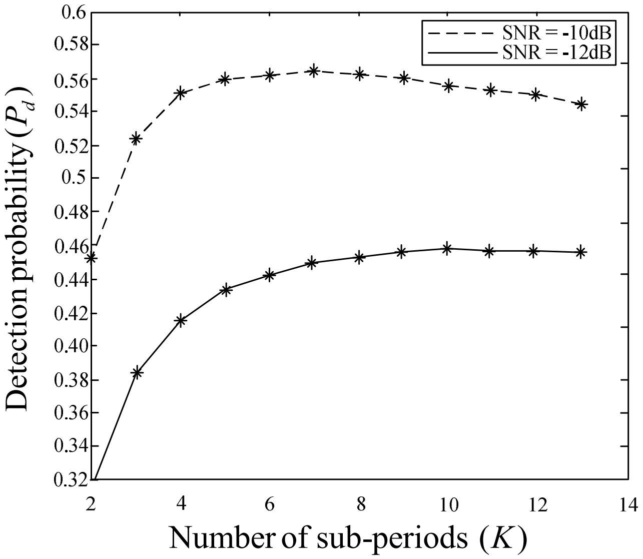


Fig. 5. The detection probability as a function of K when SNR $= -10$dB and -12dB, and $P_{fa} = 0.05$.

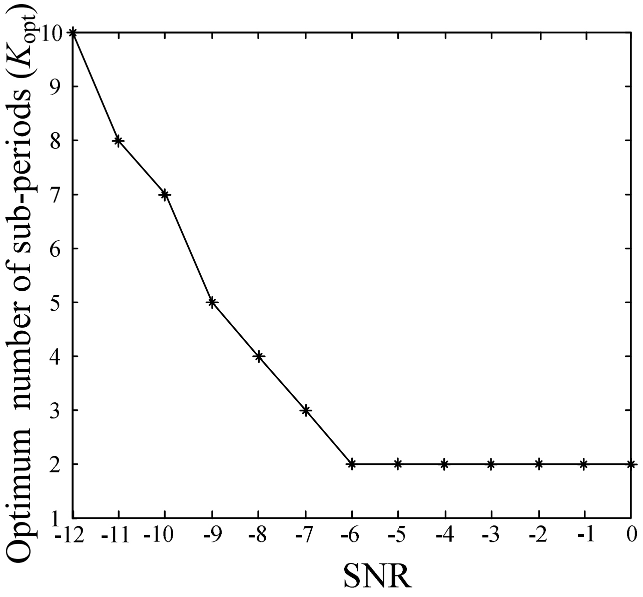


Fig. 6. The optimum number of sub-periods as a function of the Signal-to-Noise Ratio when $P_{fa} = 0.05$.

where σ_s^2 and σ_n^2 are variances of the PU signal and noise, respectively. For the scheme of [7], the product of the arrival or departure rate and the sensing period is set to 0.1.

Figure 3 shows the detection probabilities as a function of the SNR when the false alarm rate P_{fa} is 0.05 and $K = 2$, where we can clearly observe that the proposed scheme offers a detection probability improvement over the conventional schemes at low SNRs of practical interest $(-20\,\text{dB} < \text{SNR} < -5\,\text{dB})$. Although the detection probability of the proposed scheme is a little bit lower than those of the conventional schemes at high SNRs, it should be noted that all of the schemes perform well at high SNRs, and so, the difference in performance between the proposed and conventional schemes is insignificant.

Figure 4 shows the Receiver Operation Characteristic (ROC) curves of the proposed and conventional spectrum sensing schemes when the SNR = -10 dB and K = 2. From the figure, the proposed scheme is found to offer a ROC performance improvement over the conventional schemes when the false alarm probability is smaller than 0.25, which is a practical range of the false alarm probability in spectrum sensing. Specifically, it is observed that that the proposed scheme has a larger detection probability (a smaller false alarm probability) than those of the conventional schemes for a fixed false alarm (detection) probability.

The superiority of the proposed scheme over the conventional ones shown in Figure 3 and Figure 4 stems from the fact that the proposed scheme makes the final decision on the existence of the PU by giving a more weight to the sensing decision from the last sub-period unlike the conventional schemes where all of the sensing decisions from sub-periods have the same weight.

As the value of K increases, the sensing decision from the last sub-period would recognize the existence and nonexistence of the PU signal after a given sensing period more exactly; however, the reliability of the sensing decision from each sub-period would be worse, since the number of samples used during each sub-period decreases. From this observation, we can see that there exists an optimum value of K. Figure 5 shows the detection probabilities of the proposed scheme as a function of K when SNR = -12 dB and -10dB, and P_{fa} = 0.05. In this simulation, N is set to a multiple of K that is the closest to 200 when N is indivisible by K. From the figure, we can observe that 7 is the optimum value of K when SNR is -10dB, and 10 is the optimum value of K when SNR is -12dB. In addition, it is seen from Figure 6 that the optimum value K_{opt} decreases as the value of the SNR increases and eventually becomes saturated. This is because more samples are required in each sensing period to maintain the reliability of the sensing decision from each sensing period. It should be noted that the optimum number of the sub-periods is not over 10 in the SNR ranges (-12 dB $\sim$ 0 dB) of practical interest, i.e., the proposed scheme does not involve a significant increase in complexity. Thus, the proposed scheme should be applicable to real-time applications.

V. Conclusion

In this paper, we have proposed a novel spectrum sensing scheme for dynamic traffic environments. Specifically, we have first partitioned a sensing period into several sub-periods and then have performed energy detection during each sub-period. Finally, we have obtained a decision rule on the existence of a PU signal by combining the sensing decisions from sub-periods. From numerical results, we have confirmed that the proposed scheme offers a noticeable performance improvement, and also, have determined an optimum number of sub-periods.

Acknowledgment

This research was supported by the National Research Foundation (NRF) of Korea under Grants 2014R1A5A1011478 and 2012R1A2A2A01045887 with funding from the Ministry of Science, ICT&Future Planning (MSIP), Korea.

References

[1] J. Mitola, “Cognitive radio: An integrated agent architecture for software defined radio,” Ph.D. dissertation, Teleinformatics, Royal Inst. Technol. (KTH), Stockholm, Sweden, May 2000.

[2] J. Lunden, S. A. Kassam, and V. Koivunen, “Robust nonparametric cyclic correlation-based spectrum sensing for cognitive radio,” *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 38-52, Jan. 2010.

[3] F. F. Digham, M.-S. Alouini, and M. K. Simon, “On the energy detection of unknown signals over fading channels,” *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21-24, Jan. 2007.

[4] T. S. Shehata and M. El-Tanany, “A novel adaptive structure of the energy detector applied to cognitive radio networks,” in *Proc. Canadian Workshop Inform. Theory*, Ottawa, Canada, May 2009, pp. 95-98.

[5] T. Wang, Y. Chen, E. L. Hines, and B. Zhao, “Analysis of effect of primary user traffic on spectrum sensing performance,” in *Proc. Chinacom*, Xian, China, Aug. 2009, pp. 1-5.

[6] M. Luís, A. Furtado, R. Oliveira, R. Dinis, and L. Bernardo, “Towards a realistic primary users behavior in single transceiver cognitive networks,” *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 309-312, Feb. 2013.

[7] N. C. Beaulieu and Y. Chen, “Improved energy detectors for cognitive radios with randomly arriving or departing primary users,” *IEEE Signal Process. Lett.*, vol. 17, no. 10, pp. 867-870, Oct. 2010.

Utility-based Approach for Video Service Delivery Optimization

Mamadou Tourad Diallo, Frédéric Fieau
Orange Labs
Audiovisual Network Delivery
Issy les moulineaux, France
{mamadoutourad.diallo,frederic.fieau}@orange.com

Emad Abd-Elrahman
National Telecommunication Institute
Computer & Systems Depatement
Cairo, Egypt
emad.abd_elrahman@telecom-sudparis.eu

Hossam Afifi
Institut Mines-Telecom
Departement RST
Saclay, France
hossam.afifi@telecom-sudparis.eu

***Abstract*—This work aims to introduce an Utility-based approach for Video Service Delivery Optimization (U-VSDO). Through this optimization, a global utility function is calculated based on different constraints. Those constraints are considered based on separate utility function for each actor in the video service delivery chain (Content Provider (CP), Operator (OP) and Client (CL)). However, each actor has a global score for his vision, the overall optimization aims to satisfy the three actors. Our proposed methodology for this optimization is validated through different types of evaluation. First, a simulation based utility function is done for obtaining the optimal values of our optimization problem. Then, a complete GUI (Graphical User Interface) interface is built based on the main parameters for each actor. Finally, a test-bed is conducted to differentiate between two types of flows using open source Software Defined Network (SDN) controller. This part considered the standard use case for ETSI (European Telecommunications Standards Interface) in CDN (Content Delivery Network) as a Service.**

Keywords- CDN optimization; video service delivery; utility function, quality QoS/QoE.

I. INTRODUCTION

New video service delivery strategies face two challenges: pricing plan for the overall chain elements and the innovation for new added services. Many operators are struggling to maintain the Average Revenue per User (ARPU) and margins in revenues despite the high competition in market. They are searching for new optimizations that can achieve the balance between the main three actors in the chain (content providers, operators and consumers). But, the massive deployment of Over-The-Top (OTT) technology [1] is really representing a big threat for managed video services. Moreover, new opportunities brought by clients need to be studied in order to build a good utility between users needs and service requirements. Therefore, searching optimization algorithms and tools for managed video delivery networks is required.

The traditional Content Delivery Networks (CDNs) are not defined mainly for data centers virtualization but for data caching and services acceleration. Akamai is one of the most famous CDN multi providers over Internet as it handles almost 30% of global Internet traffic all over the world [2]. Hereinafter, we will explain the main challenges in video data centers in general and conduct a subjective comparison between the main actors in video service delivery.

A. Video Data Centers Issues

Online video uses a very large amount of storage in data centers and bandwidth (BW) over the Internet. In USA only, almost 50% of Internet BW is consumed by online videos [3]. Globally, one of the main issues in data centers is the movement of contents. We tried in a previous work to study the issue of content movement and video file optimization in terms of access cost from user perspective [4]. While, in another work, we focused on the QoE aspects and their effects on data retrieval or caching costs [5], the overall control performance especially in video services is still insufficient due to the main bottlenecks in data centers interconnections. Moreover, the more famous data centers over Internet proposed by Amazon [6] or Google [7] are suffering from the same problem of bottlenecks as reported in [3]. So, until the cloud solutions bring an improvement, there are still some drawbacks in content movements either within single data centers or between data centers. So, there are high incentives to search an optimized solution for big data movements and its optimization. New trends consider the Software defined network (SDN) solutions as a movement tool and enabler.

B. Comparison for the Three Actors

It is important to analyze the main actors in video service delivery chain. Then, we can describe the objectives of each actor in order to introduce his utility and the overall work motivations. Here, two comparisons are mandatory in order to build our utilities and have clear problem statements as follows:

1) Agility Comparison

The Agility is defined as the number of parameters and the ability of adaptation for the proposed system dynamically. So, the flexibility of service planning either for content adaptation or server placement is considered as an important factor in any video streaming chain. Thus, either for live streaming or VoD (Video on Demand), the easy adaptation and simple configuration of networks will enhance the overall system performance and users satisfactions at same time. Moreover, the correlation between the three actors in the video chain will lead to an optimal identification for both network capacities and users densities. Table I compares the Agility of the three actors effects in terms of some major attributes as follows:

TABLE I. AGILITY COMPARISON FOR THE THREE ACTORS

Attribute	Content Provider	Operator	Client
Video Coding	The number of layers that can be available for each content like DASH layers or HLS for mobile users	The carrier has to support different tunnels of traffic and with different rates of playing videos	The client application capacity for accommodating different coding layers and buffers required
Line Speeds	Cost consumptions for high speed deployment lines to contents hosting	Fast adaptations and scalable networks for highly on demand services	Line speed constrains either for fixed rates cost or on demand bandwidth
Capacity	Maximizing the throughputs	Minimizing the network load	Maximizing the number of clients
Quality	QoS SLA/TCA between CP & OP for an efficient content delivery with min and max thresholds of quality	Quality of service measures for adaptive bit rates	Participating in QoS/QoE reports for enhancing the overall service delivery
Devices	Hardware or Software consumed for contents virtualizations or services on demand	Dynamic allocations for resources and network virtualization to cope with on demand servers caching or placements	Device capabilities to fit with different access networks and with virtual applications

2) *Cost Comparison*

Table II gives an overall cost comparison from each actor view as follows:

TABLE II. COST COMPARISON FOR THE THREE ACTORS

Attribute	Content Provider	Operator	Client
CAPEX cost	Min cost for content adaptations	Min transmission cost for each content	Min cost for required bandwidth line
OPEX cost	Hosting servers for different layers of same content	Running cost for QoS SLA/TCA between CP & OP	Running cost for additional Bandwidth

Based on the previous two proposed comparisons and main issues in service delivery, we can formulate our problem statements as follows:

C. *Problem Statement*

We propose a global optimization utility function for each one of the three actors in the video chain. As shown in Figure 1, the three actors in the chain are in collaboration for the best service delivery. Actor 1, the content provider asks Actor 2 (the operator) to deliver some video content requested by the third Actor 3 (client). We assume that the system is real time so requests can be handled through some controller unit that manages sessions and handover decisions between CDNs based on our optimization function.

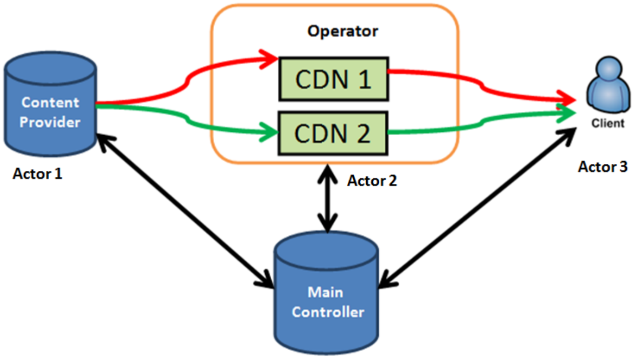


Figure 1. Main three actors in the video chain

The rest of this paper is organized as follows: Section II highlights the relevant work to this optimization and presents the different categories in video service delivery optimization. Then, Section III introduces our proposed methodology based on the new utilities constrains. The evaluation for our proposal is conducted in Section IV. Finally, this research is concluded with some future directions in Section V.

II. RELATED WORK

We can divide contributions for optimizing video delivery into three main methods: *i) the network-centric approach*, in which decisions are made at the network side (mainly by network operators), *ii) the user-centric approach* making the decision based on the user's benefit, and *iii) the context-centric approach*, where the switching decision is made by considering different context information.

i) The *network-centric*:

In this approach, decisions are made by the operators and they are principally based on their benefits.

Sylvia et al. [8] propose a distributed strategy to get network topology information, and use Internet Control Message Protocol (ICMP) ping method to measure Round-Trip Time (RTT), in order to switch to a network which has the lowest RTT. Xueying et al. [9] work on the load balancing algorithm which automatically selects network candidate based on local resource conditions. The main advantage of this method is the network resources optimization. But these techniques do not consider content provider expectations and users Quality of Experience (QoE).

ii) The user-centric:

Network switching is made in order to satisfy user’s benefits, without considering network load and content provider expectations. Ksentini et al. [10] consider Quality of

Experience measurements over different access types. After predicting a Mean Opinion Score (MOS) with Pseudo Subjective Quality Assessment (PSQA), a vertical handover (change in access network) is carried out towards the network offering the best MOS. It can be noticed that the user-centric approach has the main drawback from a load balancing perspective, since users generally consider only their own benefits while making decisions and letting the Operator and Content Provider benefits.

iii) The context-centric approach:

In this approach, the delivery decision optimization is made by considering different contexts (Content Provider, Operator, and Client). Bogdan et al. [11] propose an algorithm, called Smooth Adaptive Soft-Handover Algorithm (SASHA). Its goal is to improve the user perceived quality while roaming through heterogeneous wireless network environments. The score of each connection is evaluated based on a comprehensive Quality of Multimedia Streaming (QMS) including the following metrics: QoS (Quality of Service), QoE (Quality of Experience), Cost, Power efficiency and user preferences. The idea is to adapt delivery in the network that has the best (QMS) score. The disadvantage is the no consideration of content provider expectations in the adaptation process.

Suciu et al. [12] propose Hierarchical and Distributed Handover (HDHO) method, a distributed handover decision framework which takes into account the objective of Content Provider by considering the content requirements in terms of resources, Operator in terms of network load and user preferences by considering cost sensibility. Even if, this proposal takes into account the aim of each actor on the delivery chain, some relevant parameters are omitted. In content provider side the cost of transmitting the content in a network is missed, in network side the cost and hardware status are absent, in client side the perceived quality of experience is not taken into account.

In order to maximize a perceived quality of experience in users' side, respect conditions of content providers and the operators' benefits, we need to define a new video delivery optimization which takes into account the objective of each actor. Moreover, in such a dynamic environment composed by different devices with different characteristics, variables network conditions with different cost/load and content providers with different expectations, we propose the U-VSDO algorithm that handles all those parameters as explained in the next section.

III. Proposed Methodology

The purpose of this section is to explain the steps of the optimization approach which takes into account the objective of Content Provider (CP), Operator (OP) and the User. Our approach is based on the definition of three entities, each with their goals as follows:

- The objective of Content Provider is to send the Content in the network with a minimum cost and still manage the Content expectations in terms of requirements (for example the minimum required throughput for the content).
- The objective of the Operator is to transmit content on its network (CDN1 or CDN2 in our example) while keeping the load as lower as possible.
- The objective of the client is to improve the Quality of Experience besides the QoS.

The Utility-based Video Service Delivery Optimization (U-VSDO) will take into account the goals of each actor in addition to the main constrains. As shown in Figure 1, the optimization decision will be managed by the Main Controller after solving the optimization problem. This controller can be for example an SDN controller as will be explained in Section IV for SDN Network Function Virtualization NFV [13]. So, we can solve the problem by the following steps:

A. Problem Formulation

We used the utility functions to calculate the scores of each actor; this is very useful to characterize the satisfaction derived from a parameter.

The function must have the following characteristics:

- The function increases with parameter x and has a maximum of 1,
- When x is "low", the function tends to zero.
- The possibility to have normalized results between [0, 1].

Several functions meet these criteria. Moreover, we decided to use the utility function: (1- e^{-x}), as the work in [12] [14], where x is a parameter of the function. In future work, we will further investigate the influence of others utility functions in our optimization problem.

Hereinafter, we introduce the details of each actor utility function based on the previous propositions either for utility type or normalization way. Then, a global score utility will be calculated under the main constrains defined for each actor as follows.

As the work in [12], we have two types of parameters:

- The positives parameters: High values are better, example (throughput, available hardware, etc.), then for an utility function we took the parameter directly.
- The negatives parameters: Low parameters are better, example (cost, network load, etc., then for these parameters we choose $\frac{1}{\mathrm{cost}}$ for example.

- ***<u>For Content Provider:</u>***

$$S_{cp}(i,j) = (1 - e^{-\frac{1}{C_{cp(j)}}} + 1 - e^{-Dr(i)}) \,.\, C_{cps} \,.\, Ds \qquad (1)$$

$$S_{cp}(i,j) = (2 - e^{-\frac{1}{C_{cp(j)}}} - e^{-Dr(i)}) \,.\, C_{cps} \,.\, Ds \qquad (1)$$

where:

- ○ $S_{cp}(i,j)$, is the score related to Content Provider (CP) for flow j in network i.
- ○ C_{cp} = UNIT cost per Mbyte, is the cost of transmitting the content in the network (CDN1 or CDN2) in our example).
- ○ C_{cps} = (C_{cpmax}(j),$C_{cp}(j)$) = 0, when $C_{cpmax} < C_{cp}$, S_{cp}= 0.
- ○ C_{cpmax}, is the maximum cost that the content provider is ready to pay.
- ○ D_r, is the available throughput.
- ○ *Ds = (D_{ref} (j) ,D_r (i))* =0, when $D_{ref} < D_r$
- ○ D_{ref}, is the required video throughput.

Note that: (A, B) means that; when A < B then (A,B) =0 and 1 otherwise.

- ***For Operator:***

$$S_{op}(i) = (3 - e^{-\frac{1}{Cop\,(i)}} - e^{-\frac{1}{NL(i)}} - e^{-H(i)}).NLs.Cops.Hs \quad (2)$$

where:

- ○ $S_{op}(i)$, is the score related to Operator in network i.
- ○ $Cop = Opex + Capex$; is the cost from the operator side.
- ○ NL , is the network load.
- ○ *NLs=($NL_{max}(i)$, NL (i)) =0, when $NL_{max} < NL$*
- ○ NL_{max}, is the maximum acceptable network load.
- ○ *Cops= ($Cop_{max}(i)$,Cop(i))=0, when Cop_{max}< Cop*
- ○ Cop_{max}, is the maximum price that the operator is ready to invest.
- ○ *H* is the required hardware threshold.
- ○ *Hs= (H(i), H_{min}(i)) = 0, when $H < H_{min}$*
- ○ H_{min}, is the minimum required hardware for considered service.

- ***For Client :***

$$S_{Cl}\ (i,j) = \frac{MOS_{NET}\ (i,j)}{S_{max}} \qquad (3)$$

where:

- ○ $MOS_{NET}\ (i,j)$ corresponds to the satisfaction obtained by users in network i for flow j. It is a parametric model which computes the Quality of Experience function of contexts information (environment) [15], the model takes into account parameters such as the device type, the video content type and the quality of the network link in order to predict the Quality of Experience. The analytical function is called MOS_{NET} and is presented in the equation below:

$$MOS_{NET}(i,j) = A + B * e^{-C*\frac{Dv(j)}{Dr(i)}} \quad (4)$$

- ○ *A, B* and *C:* are the model parameters calculated by using subjective test data from different experiments.
- ○ S_{max} , is the maximum value of MOS_{NET} which correspond to the normalized factor

So, the general optimization problem can be formulated as follows by total score:

$$S_T = \alpha * S_{cp} + \beta * S_{op} + \mu * S_{Cl} \quad (5)$$

where : α, β, μ are the weights of entities in the global optimization and : $\boldsymbol{\alpha + \beta + \mu = 1}$

The weighting parameters define the importance of each actor in the optimization decision. In our work we decided that the Content Provider, the Operator and Users have the same weight, then: $\boldsymbol{\alpha = \beta = \mu = \frac{1}{3}}$

B. Optimization Problem Constraints

In this section, we summarize the main utility functions for the computed scores and their constraints that will be implemented in the next section and appeared in the GUI interface as follows:

$$Content\ Provider : S_{cp}(i,j) = 2 - e^{-\frac{1}{Ccp(j)}} - e^{-Dr(i)}$$

$$Operator : \quad S_{op}(i) = 3 - e^{-\frac{1}{Cop\,(i)}} - e^{-\frac{1}{NL(i)}} - e^{-H(i)}$$

$$Client : \qquad S_{cl}(i,j) = \frac{Aj + Bj * e^{-Cj*\frac{Dv(i)}{Dr(j)}}}{S_{max},j}$$

Objective: $\boldsymbol{maximize\ (\alpha * S_{cp} + \beta * S_{op} + \mu * S_{cl})}$

$$\begin{aligned} \textit{Subject to}: \quad & C_{cp} < C_{cpmax} \\ & C_{op} < C_{opmax} \\ & Dr < Dref \\ & NL < NL_{max} \\ & H_{min} < H \end{aligned}$$

IV. IMPLEMENTATION AND EVALUATION

To validate our work, we propose two ways. First, we are going to optimize the utility function parameters through a simulation tools using Matlab. Then, the decision output of this optimization will take the form of graphical interface for doing many scenarios. Second, we will do a testbed for decision making based software defined network (SDN) solution to differentiate between CDNs caching. This solution conforms to the ETSI solution use case 8 for virtual CDN-as-a-service [13].

A. Validation Based Simulation

The validation based simulation has been conducted based on some real test captured from last championship Roland Garros (RG) [16]. Roland Garros is a major tennis tournament held over two weeks between late May and early

June 2013 at the Stade Roland Garros in Paris, France. Some analysis has been proposed in our previous work [17]. In this work, we studied and analyzed users engagement during this event based on some real time measures done based on the Orange platform. Then, we will extend the study and the analysis to conform to the three actors in video. Samples from the RG tests are shown in the following Table III. Parameters such as buffering time, startup time and playing time are considered.

TABLE III. SAMPLES FROM ROLAND GARROS MEASURES

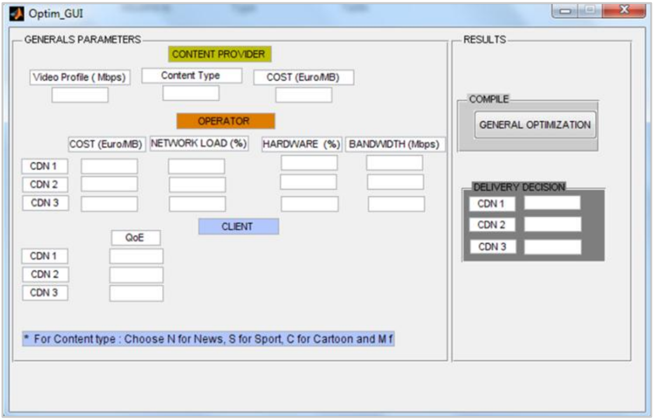
Buffering ratio (%)	Startup time (Seconds)	Average encoding (kbps)	Playing time (seconds)
0,84375	2,671875	807,0869565	446,85
1,1761176	0,8860886	973,3505747	3600,33
0,785625	0	970,070922	1521,11
1,15	0	963,8943089	1356,33
2,734375	0,59375	2040,769811	2791,39
0,682039	0,6560375	945,2185792	1938,35
0,5624928	0,8281144	955,7608696	455,50

We implemented a complete Graphical Tool (GT) to be used by the operators in their networks design and optimization. This graphical tool is built based on Matlab code.

Figure 2 illustrates the main construction steps as divided into two parts:

- Creating general parameters: which means defining the basic topology elements and factors in the three actors (CP, OP and CL) i.e. the main profiles for each video and CDN.
- Calculating results: Calculating the general score for all actors and show the selected CDN as best path for video profile.

Actually, we simulate the global utility function and calculate the scores for different networks for our approach U-VSDO. Moreover, and in order to facilitate the decision making output by each operator running our methodology, we developed a GUI interface to cope with the three utility functions for the three actors main parameters as shown in Figure 2.



* For Content type : Choose N for News, S for Sport, C for Cartoon and M f

Figure 2. GUI interface for U-VSDO approach

After finishing this simulation, we conducted a brief comparison between our approach U-VSDO and other similar techniques that used utility functions for decision making based multimedia handover like SASHA [11] and HDHO [12]. The results indicated in Table IV highlighted the main parameters considered as supplementary by our approach U-VSDO over other ways.

TABLE IV. COMPARISON BETWEEN U-VSDO TO OTHER APPROACHES

	SASHA	HDHO	U-VSDO
OP cost	x	x	√
CP cost	√	√	√
Content Type	x	x	√
Device Type	x	x	√
Client Type	x	x	√
Network Load	√	√	√
H/W Status	x	x	√

Finally, Figure 3 represents the correlation between this work and our previous work [17] for different types of media tested under our approach. Figure 3 handled different types of videos (News, Sport, Music and Animation) in terms of which CDN can achieve high scores in order to satisfy user engagement and all actors' satisfactions for our methodology.

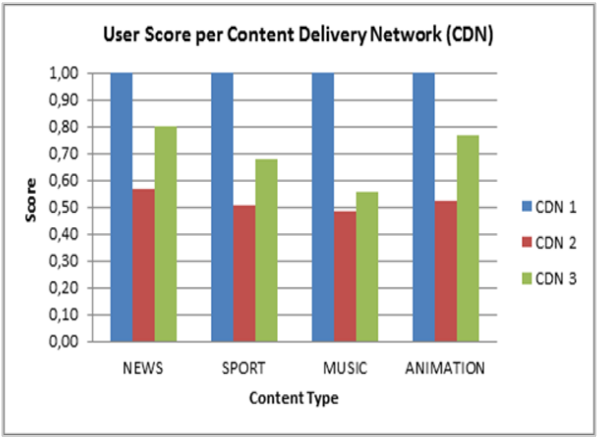


Figure 3. User score for different types of media to different CDNs

B. Validation Based Testbed

In this part, we try to validate our approach using software defined network controller through the SDN implementations based testbed. The testbed architecture is shown in Figure 4. We simulate the traffic source as video on demand servers by VOD1 & VOD2 (using VLC Servers for same contents but on different format: one Standard Definition (SD) and the other High Definition (HD)) and the client will also use a VLC client.

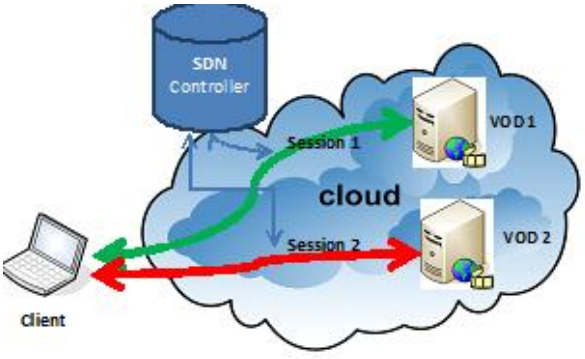


Figure 4. Proposed testbed architecture

The experiment for simulating SDN controller is done using Mininet open source package [18]. Using this package, we can build a complete architecture of virtual topology (including VM clients, VM servers, Virtual Open switch and the session controller based an OpenFlow [19]).

As SDN offers Networking-as-a-Service (NaaS) and Network Function-as-a-Service (NFaaS), the main objective is to measure the performance in case of obtaining the contents from VOD1 (SD) and due to the QoE index; the session will be transferred to VOD2 (HD) for same contents. Also, we can see that this is a type of session based hijacking using SDN controller. The response time for sessions hijacking is calculated for different types of video bit rates as shown in Figure 5. As shown in this figure, the switching time is acceptable for different types of videos and is conform with the ITU recommendations for quality of streaming.

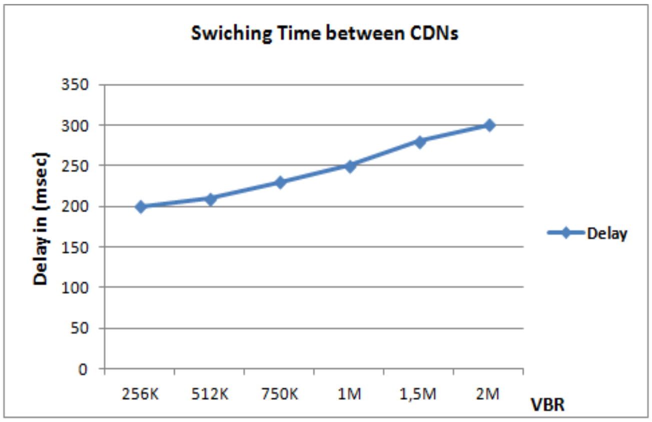


Figure 5. The response time for sessions hijacking against VBR (video bit rate) as measured during the experiment.

V. Conclusion

An optimization mechanism is presented and evaluated. It solves the utility function optimization for the three common actors in video streaming chain (CP, OP and CL) for any data centers, including their roles and objectives in video chain. The proposed methodology U-VSDO is evaluated by two ways. First, through a simulation for global utility function and our approach gave good results compared to other methods like HDHO or SASHA in terms of value added parameters in decision making. Secondly, a testbed is evaluated to validate the CDN-as-a-Service controlled by an SDN controller and the consumed time for sessions hijacking is measured for different video bitrates.

In the next work directions, we will consider the real-time video service optimization based on adaptive technique suitable for the real-time nature.

References

[1] Over The Top (OTT) : http://www.pace.com/Documents/Investors/Presentations/100609_TechnologyBriefing.pdf, last visit: May 2014.

[2] Akamai: http://www.akamai.com/ , last visit : May 2014.

[3] Fierce online video Report: http://www.fierceonlinevideo.com/special-reports/moving-big-data-video-transport-providers-keep-sporting-events-hollywood-pa#ixzz2nBe1xrds: December 11, 2013.

[4] Abd-Elrahman E. and Afifi H., “Optimization of File Allocation for Video Sharing Servers”, NTMS 3rd IEEE International Conference on New Technologies, Mobility and Security (NTMS),Cairo-Egypt , 20-23, pp.1-5, Dec. 2009.

[5] Abd-Elrahman E., Rekik T., and Afifi H., ‘‘Optimization of Quality of Experience through File Duplication in Video Sharing Servers”, EuroITV2012, UP-TO-US workshop, Germany: July 2012. pp.286-291

[6] Amazon Web Services: http://aws.amazon.com/, last visit: July 2014.

[7] Google: http://www.google.com, last visit: Augest 2014.

[8] Sylvia R., Marc H., Richard Maning K., and Scott S., “Topologically-Aware Overlay Construction and Server Selection” INFOCOM 2002, pp. 1190-1199.

[9] Xueying J., Shiyao L., and Yang Y., “Research of Load Balance Algorithm Based on Resource Status for Streaming Media Transmission Network” CECNet, 2013,pp.503-507

[10] Ksentini A., Viho C., and Bonnin JM., “QoE-aware Vertical Handover in Wireless Heterogeneous Networks”, IEEE, 2011, pp.95-100.

[11] Bogdan C. and Gabriel M., “A Quality-Oriented Handover Algorithm for Multimedia Content Delivery to Mobile Users” Broadcasting IEEE Transactions on, pp. 437-450

[12] Suciu LL. and Guillouard K., “A Hierarchical and Distributed Handover Management Approach for Heterogeneous Networking Environments”, ICNS 2007, vol., no, pp.77,77, 19-25 June 2007.

[13] ETSI GS NFV 001 v1.1.1 ,“Network Functions Virtualization NFV; Use Cases”, October 2013.

[14] Penhoat J., Guillouard K., Lemlouma T., and Mikaël S., “Analysis of the Implementation of Utility Functions to Define an Optimal Partition of a Multicast Group”, ICN 2011, The Tenth International Conference on Networks. 2011.

[15] Diallo M.T., Marechal N., and Afifi H., “A Hybrid Contextual User Perception Model for Streamed Video Quality Assessment”, Multimedia (ISM), 2013 IEEE International Symposium on , vol., no., pp.518,519, 9-11 Dec. 2013.

[16] Roland Garros: http://www.rolandgarros.com/, last visit: February 2014.

[17] Diallo M.T., Fieau F., and Hennequin J.B., “Impacts of Video Quality of Experience on User Engagement in a Live Event” ICME EMSA 2014.

[18] Mininet: http://mininet.org/, last visit: January 2014.

[19] https://www.opennetworking.org/, last visit: January 2014.

Perceptual Semantics for Video in Situation Awareness

María Alejandra Pimentel-Niño, María Ángeles Vázquez-Castro and Iñigo Hernáez-Corres

Dept. of Telecommunications and Systems Engineering
Universitat Autònoma de Barcelona
Bellaterra, Spain 08193
Email: {mariaalejandra.pimentel, angeles.vazquez}@uab.es, inigo.hernaez@e-campus.uab.cat

***Abstract*—We introduce novel perceptual semantics for video adaptation in multimedia communications. The target is to enhance situation awareness in non-computer aided processes as in emergency operations. Our proposed perceptual semantics relate to end user requested resolution in the temporal domain for a better assessment of event's evolutions seen from streaming video. Adaptation is enabled at transmission via a perceptual semantics feedback loop to adapt source coding on-the fly in terms of frame rate. The overall framework contemplates the use of an underlying cross-layer optimization that copes with network congestion and erasures in best effort scenarios. We show through simulations that within the proposed framework, the perceptual semantics are preserved. Moreover, we show it complies with information-centric-networking philosophy and architecture, such that it is in line with content-aware trends in networking.**

Keywords–Perceptual semantics; adaptive video; situation awareness; QoE.

I. INTRODUCTION

The motivation of the work presented here is the possibility of enhancing situational awareness using video streaming through constraint networks. We target non-computer-aided scenarios, where there is no artificial intelligence behind interpreting sensory information from the received video.

We consider band-limited, wireless best effort networks, as possible means of communications for point-to-point live video streaming during scenarios such as in emergency operations. Such networks pose a number of constraints affecting Quality of Service (QoS), namely, congestion and erasures. The topology envisioned is that of live user-generated content being upstreamed to proper decision-makers.

Transmission alternatives that cope with the aforementioned network constraints include UDP- based frameworks for live or real-time applications, with quality based [1], Quality of Experience (QoE) driven cross-layer optimization [2], or TCP-friendly [3] adaptive algorithms. Dynamic Adaptive Streaming for HTTP over TCP streaming is suitable for server-client architectures and video on demand applications, but with degraded performance in lossy networks with large propagation delays [4]. With regards to the lossy nature of the network, forward erasure correction methods are able to provide enough protection and maintain QoE [5][3][6].

While these transmission frameworks can be satisfactory to guarantee QoS/QoE in video-for-entertainmet scenarios, we propose an additional dimension to target specific user demands in scenarios using video for other purposes, such as in emergency operations. We propose to improve non-computer-aided situation awareness beyond standard improvements by means of perceptual semantics.

In multimedia, "classic" semantics deals with heterogeneous metadata that sensors observe and/or tag when capturing video. As such, it has applications in information retrieval, integration and aggregation of varied data types as in semantic-aware delivery of multimedia [7]. Further, semantic tagging describing pure observations is used in computer-based systems with artificial intelligence to perceive and abstract situations [8]. Rather than doing perception through classic semantics, we propose a novel human-analysis-driven perceptual semantics to tag the videos, based on the temporal/spatial characteristics a user is perceiving as means to improve situation awareness.

The term perceptual semantics has been used by Cavallaro and Winkler [9] for automatic feature extraction of video based on image segmentation and target internal changes within the video source coding mechanisms. Our approach differs, as it does not limit to particular feature extraction and it focuses on offering a solution for video communications in a non-intrusive manner towards video codecs. Further, we involve the user in tagging first-level perceptual features, for perceptual-based networking, rather than use only observations as in [10]. To the best of our knowledge, such diversion from classic semantics has not been explored before.

Finally, we frame the novelty of perceptual semantics such that it complements cross-layer optimization schemes that help cope with network constraints. Within this framework we enable a perceptual semantic adaptation loop, which will target the specific user's demand for improved perception, comprehension and further projection in situation awareness processes. Additionally, this framework can be mapped to current content-centric approaches of information-centric-networking.

The structure of the paper is as follows. We present the perceptual semantics model in Section II, followed by the integration of perceptual semantics within a video adaptation framework in Section III. We show simulation results in Section IV, and draw final conclusions in Section V.

II. PERCEPTUAL SEMANTICS MODEL

In this section, we derive the model for perceptual semantics in the context of situation awareness and how we propose to perform semantic tagging.

A. Spatial/temporal decoupling for situation awareness

Situation awareness enables good decision-making [11] and hence, it is a major asset in, e.g., emergency operations. A

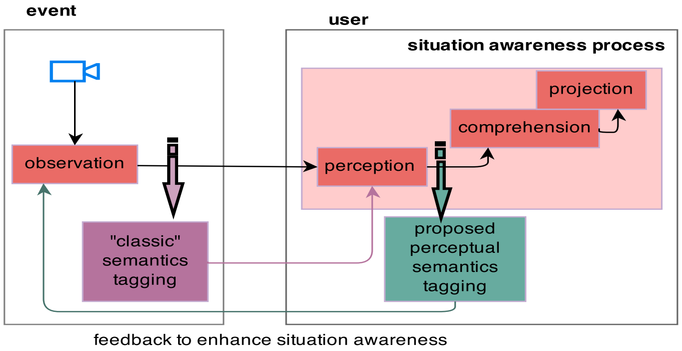

event
user
situation awareness process
projection
comprehension
observation
perception
"classic" semantics tagging
proposed perceptual semantics tagging
feedback to enhance situation awareness

Figure 1. Perceptual semantics vs “classic semantics”

broadly accepted definition is “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.” [12]. The three-level model is thus inferred, namely: perception, comprehension, and projection.

We focus on the spatial and temporal advantages of video as a source of information for situation awareness. First, the possibility of capturing dynamic scenes improves the assessment of temporary evolving events. Second, video can provide visual spatial accurate accounts of an ongoing situation [13].

If the temporal and spatial perceptual characteristics of the video satisfy the situation-dependent specific user resolution, then the user satisfaction will be fulfilled, and further higher level cognitive processes will be benefited.

B. Semantic tagging

Based on the spatio/temporal identification of perceptual features, our proposal is to utilize the end-user’s (analyst) perception, to do semantic tagging that enables an enhancement of the received video stream signal tailored to the user’s demand.

Semantic tagging is hence performed to describe perceptual features in the video and as such represents more complex abstractions of a viewed scene. In comparison, classic semantics tagging would focus on unprocessed sensorial observations [8]. The difference between both approaches in semantics is shown in Figure 1.

In scenarios where perception is not achieved by artificial intelligence, it is the human analysis that will interpret the sensory information and follow the three steps in the situation awareness model. Hence, the semantic tagging is performed by the user, as he is ultimately the one perceiving and foreseeing what might be of interest in the video.

We propose a tagging that would indicate the temporal/spatial predominance according to the level of perception of the user. A tag indicating predominance of temporal features, means the user is perceiving a situation that demands more attention to the dynamics of the scene (e.g., rapid movements, evolution of an environmental hazard). On the other hand, a predominance of spatial features indicate moments of less movement but densely overloaded frames that requires more detail to identify features (e.g., identifying persons or details in a emergency scene).

III. Integration of perceptual semantics to video adaptation

We propose to integrate the perceptual semantics with video adaptation in order to provide to the user the required perceptual level for situation awareness. Further, we propose to map the tags to actions, such that the specific perceptual features are enhanced.

Following, we describe how our perceptual semantics model can be mapped to video coding characteristics and propose an algorithm to meet the end-user’s demands. We comment on protocol aspects in the implementation and propose an integrated framework with an adaptive video solution.

A. Mapping

We focus on using our proposed perceptual semantics for enhancement at source coding level. In single layer or scalable layer video encoding of state-of-the-art codecs, three types of resolution are defined, namely temporal (frame rate), amplitude (quantization step), and spatial (frame size).

We map enhancement of temporal features to higher frame rates, and predominance of spatial features to higher spatial and amplitude frame resolution. In this way, dynamics of the scene can be more closely followed (temporal preference) and details of a scene can be better identified (spatial preference). The mapping is intuitive and relies on the intrinsic architecture of video codecs currently in use in a non-intrusive manner, to facilitate the video communications. Finally, we show an example architecture within Information-Centric Networking (ICN) networking of a typical emergency scenario.

B. Algorithm

We propose to map the perceptual semantics to a system quantified with the variable $\alpha \in [0, 1]$. $\alpha = 0$ and $\alpha = 1$ express full preference of the spatial and temporal perceptual features, respectively. Intermediate values of α represent weighed combinations of spatial and temporal preferences.

We denote the feasible set of finite values of frame rate, as $F_T(r_{APP})$, while $F_S(r_{APP})$ is the feasible set for the spatial factors, both a function of video coding rate r_{APP}. Note that higher frame rates and frame sizes are possible to attain with higher r_{APP} [14], hence the feasible sets $F_S(r_{APP})$ and $F_T(r_{APP})$ corresponding to higher values of r_{APP} will contain more number of possible values that can be chosen from. For example, in the case of scalable video coding, if temporal dyadic scalability is performed, the available values of frame rate contained in $F_T(r_{APP})$ would be the base layer frame rate and the frame rates from enhancement layers that would add up to i.e. a full 30Hz frame rate if r_{APP} is sufficient: $F_T(r_{APP}) = \{$3.75Hz, 7.5Hz,15Hz,30Hz$\}$.

In order to choose the appropriate value of frame rate and resolution according to our mapping of perceptual semantics, we formulate the following optimization function:

$$\begin{aligned} (r^*_{fr}, s^*_{fr}) &= \max\ (\alpha \bar{r}_{fr} + (1-\alpha)\bar{s}_{fr}) \qquad (1)\\ s.t. &\quad r_{fr} \in F_T(r_{APP}) \text{ and } s_{fr} \in F_S(r_{APP}) \end{aligned}$$

where $\bar{r}_{fr} = {}^{r_{fr}}/_{r^{max}_{fr}}$, and $\bar{s}_{fr} = {}^{s_{fr}}/_{s^{max}_{fr}}$ are the normalized values of frame rate r_{fr} and spatial/amplitude resolution

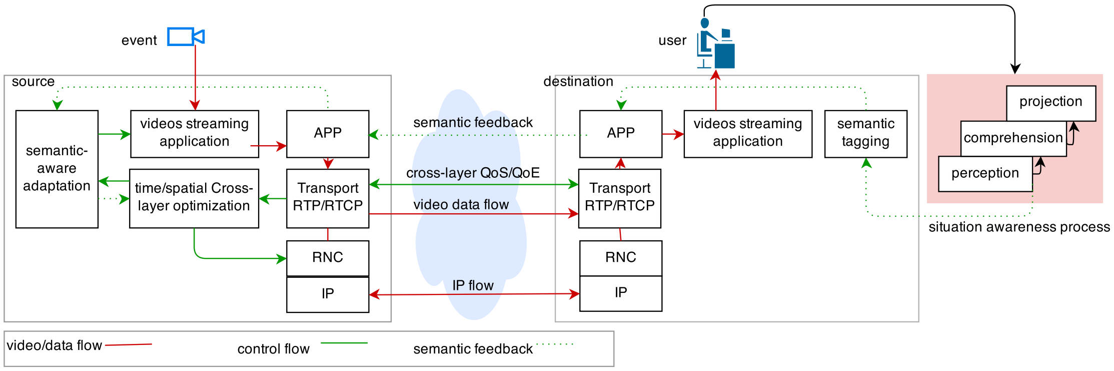


Figure 2. Block diagram proposed cross-layer framework and APP-to-APP perceptual semantics loop

s_{fr} with respect to maximum available values set for the application. Note that the optimization in (1) can be applied to single layer video coding or scalable video coding.

C. Implementation and compliance with standards

Figure 2 shows the diagram of the proposed framework.

1) Cross-layer framework: We assume an underlying standard cross-layer framework, that uses transport layer feedback and interacts at transport-APP layers and transport-network layers of the IP protocol stack.This framework provides the application layer rate r_{APP} that can be used by the codec (such that the video coding rate equals the application layer rate), for an on-the-fly adaptive video subject to network constraints. Moreover, it provides to the network layer the necessary parameters to perform forward erasure protection. The cross-layer optimization is handling feedback with the standard RTP/RTCP protocol [15] (Real Time Protocol/Real Time Control Protocol). Note that we have assumed Forward erasure protection being performed at network layer, in particular using Random linear Network Coding (RNC).

The cross-layer optimization has been designed such that it copes with the network impairments that directly affect negatively the spatial/temporal aspects of video and is therefore QoE-driven. This time/space cross-layer optimization associates congestion with temporal impairments in video playback such as freezes. In addition, it associates erasures with artifacts degrading video quality. Further, we keep in mind that higher video quality is achieved with higher video codec rate, r_{APP}.

The above assumption is relevant in the design, given that the cross-layer optimization is able to mitigate the negative effects of network degradations due to congestion and erasures. Therefore, network degradations will minimally affect the job of the perceptual semantics when enhancing the temporal and spatial features necessary for situation awareness. We will further discuss this with the numerical results in Section IV.

2) Perceptual semantics: Figure 2 further shows how the cross-layer optimization is integrated with the perceptual semantics loop. The video streaming application uses a state-of-the-art codec such that the frame rate, frame size and codec rate can be configured on-the-fly, either as a single layer or a scalable layer coding.

In order to facilitate the perceptual semantics role, we use a return path to send the tags chosen by the user according to the perceptual semantics. The semantics-aware adaptation block in Figure 2 interprets the semantic tags coming from the end-user by mapping it to the proper decisions and forwarding to the video codec, as explained in Section III.

Following the trends in current network architectures, we propose to use semantic web protocols to enable the APP-to-APP cross talk of the semantic tagging [10]. At the transport layer, the application-specific information can be encapsulated into RTCP feedback packets compliant with the extended reports defined in RFC4585. This way, the perceptual semantics feedback loop is coherent with the cross-layer optimization.

3) Coherence with ICN networking: Considering future architectures, our framework complies with a semantic information-based network [16]. The aim of the Information-Centric Networking (ICN) approach is to integrate content delivery as a native network feature, where focus is not on the network as an enabler of communication links but as a platform for information dissemination. ICN could allow for future enhancements to the perceptual semantics as proposed in this paper. In particular, our approach is coherent to the receiver-driven nature of ICN. Further, caching, one of the appealing attributes of ICN in data delivery, could enable more actions at intermediate nodes concerning the incoming video stream. The philosophy of ICN by which content information is available to network/forwarding layers will allow the semantic loop we have created to trigger further actions at these intermediate nodes, such as adaptive network coding to enhance last-mile network reliability.

Figure 3 shows the topology of our framework mapped to the publish/subscribe ICN architecture for live streaming by Tsilopoulos et al. [17]. It is a typical example of an emergency application over a satellite access network, whose gateway can be mapped to the *rendevouz* (RN) and *topology manager* (TM) nodes. The publisher (operator in the ground during an emergency) announces that it has a publication available to the RN node. The subscriber (end-user/decision maker) issues a subscription, as he is interested in obtaining live feed of the current on-going events of the emergency. The RN and TM nodes find the publisher and resolve the publisher/subscriber path. The subscriber can issue petitions or unsubscribe, and in our framework, issue perceptual semantics tagging, which the publisher will receive through the RN nodes.

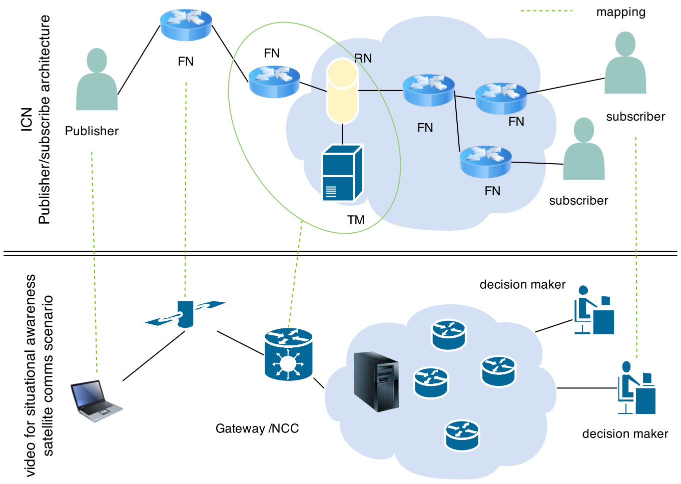


Figure 3. Publish/Subscribe architecture suitable for our proposed perceptual semantics framework

TABLE I. FEASIBLE SETS CONSIDERED FOR SIMULATION

r_{APP} (in kbps)	Feasible Set F_T	Feasible set F_S
$r_{APP} \leq 64$	{3.75, 7.5,10,15}	{QCIF}
$64 < r_{APP} \leq 192$	{3.75, 7.5,10,15}	{QCIF,CIF}
$192 < r_{APP} \leq 384$	{3.75, 7.5,10,15}	{CIF,QCIF}
$384 < r_{APP} \leq 500$	{3.75, 7.5,10,15,30}	{QCIF,CIF,640x360}

IV. SIMULATION RESULTS

We simulate a realistic scenario typical of emergency operations, where mobile satellite services are used to upstream live video from field, to proper decision makers remotely located. To our knowledge, there is no similar framework in the literature to match our proposed perceptual semantics framework and hence comparison to solutions that do not have such aim would be unfair. Therefore, our results are compared to not having such kind of framework.

A. Setup

We use a simulation system that allows to test the proposed framework shown in Figure 2. The video streaming application is simulated by generating packets of size l encoded at a rate r_{APP} and frame rate r_{frame}.

1) Cross-layer optimization setup for congestion and erasures: This block receives as inputs the feedback from the receiver on current network conditions, and outputs the rate r_{APP} that the video streaming application is allowed to use, and the code rate ρ to be used for erasure correction, such that the transmission rate is $R = {}^{r_{APP}}/_{\rho}$.

The transmission rate is online optimized through a QoE delay-driven optimization at the sender side that uses receiver feedback, as the one proposed in [2]. The resulting discrete rate control update is given by (2)

$$R(t_{k+1}) = R(t_k) + f(\tau(t_k - \tau_D)) \qquad (2)$$

where $f(\cdot)$ is a function of the delay τ measured at time $t_k - \tau_D$, a delayed value due to the propagation delay τ_D in the feedback loop. Updates on network measurements are received every $T_{samp} = t_{k+1} - t_k$ seconds.

Further, our additional novelty to cope with erasures, is the use of adaptive network coding with Systematic Random linear Network Coding, (SRNC). We use SRNC due to similar performance to optimal forward erasure correction codes like Reed-Solomon [5], but higher flexibility and compliance with future network-coded networks. For a rate budget given by R in (2), the code rate $\rho = r_{APP}/R$, chosen for SRNC is maximized such that the performance meets a target residual erasure rate given the current erasure rate ϵ of the network. Hence the application layer rate r_{APP} is maximized.

2) Network simulation: We simulate a network as a FIFO finite queue of available rate r_{av} with erasure rate ϵ. Simulated packets are transmitted at the obtained rate R.

SRNC uses the allocated code rate ρ to meet the complete budget rate R, such that its performance meets the residual erasure rate ϵ^{res}.

Congestion events are simulated as a drop (step-like) in maximum available rate r_{av}^{max} to r_{av}^{min} that occurs halfway through one streaming session, at $T/2$ such that $\eta = \frac{r_{av}^{max} - r_{av}^{min}}{r_{av}^{max}}$, with $\eta \in (0, 1]$. (Higher η means higher congestion). Each simulation, corresponding to one streaming session, lasts 300s, one corresponding value of η and ϵ.

The values used correspond to a realistic satellite network commonly used in emergency operation, operating in the L-band offering up to 500kbps uplink in best effort mode.

3) Perceptual semantics: We model the user's semantic tagging from temporal/spatial features with the parameter α. α may vary over time throughout one single streaming session, such that the sender is receiving feedback of this changes and will adapt to them using, e.g., (1). We assume these tags are changed by the user with a period of at least 10s. Three cases are considered for variation of semantic tagging, namely, TAG_T: only temporal tagging for the entire session, TAG_S: only spatial tagging, TAG_{TS}: alternating tags, each of 10s.

Table I summarizes the feasible sets for values of frame rate dependent on r_{APP}, in order to solve the algorithm in (1). The values chosen correspond to typical feasible combinations in current state-of-the art codecs.

B. Metrics

The following metrics relate to the effects of the network constraints in terms of Quality of Experience.

1) QoE_A. : This metric is related to degradations due to erasures in the network, that cause artifacts in the image: $QoE_A = 1 - \bar{p}$, where $\bar{p}$ is the average packet loss rate at the receiver. $QoE_A \in [0, 1]$.

2) QoE_F. : This metric is related to degradation due to congestion, that cause freezes in video playback. $QoE_F = 1 - \bar{f}$ where $\bar{f}$ is the probability of freezes occurring in the playback. A freeze is the event where the time elapsing between two consecutive frames displayed exceeds a tolerated threshold. $QoE_F \in [0, 1]$.

3) $\hat{\alpha}$ and Δ_α.: These metrics are related to the performance of the adaptation through perceptual semantics. We measure the value achieved by the algorithm as $\hat{\alpha}$, and the mean absolute error with respect to the user's demanded α, as $\Delta_\alpha = |\hat{\alpha} - \alpha|$.

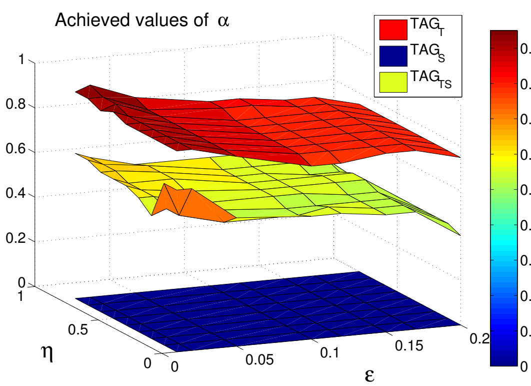


Figure 4. Achieved values of α vs. η vs. ϵ, using cross-layer optimization

4) Ω.: Combined metric to measure tradeoffs of using perceptual semantics with and without cross-layer optimization. It is defined as:

$$\Omega = w_1{\cdot}QoE_A + w_2{\cdot}QoE_F + w_3{\cdot}(1 - \Delta_\alpha)$$

with $w_1 + w_2 + w_3 = 1$. $\Omega \in [0, 1]$. The best performance, i.e., $\Omega = 1$, occurs when no losses degrade the video ($QoE_A \rightarrow 1$), freezes in playback are minimal ($QoE_F \rightarrow 1$) and the perceptual semantic adaptation matches the one requested by the user (Δ_α).

C. Results

1) Perceptual semantics with and without cross-layer optimization: Figure 4 shows the performance with respect to metric α of the perceptual semantics together with the cross-layer optimization, as a function of congestion drops, η, and erasures ϵ. Each surface corresponds to one of the three cases of time varying semantic tagging. TAG_T achieves high values of α close to the tagged from the user, representative of preference on temporal features, while TAG_{TS} offers an intermediate values, corresponding to the alternating tags. α only reflects on the performance of the perceptual semantics algorithm, and whether it is capable to achieve the expected user demand. However, it does not reflect the effects on QoE_F and QoE_A, directly affected by network degradations. Ω will express the full performance as a whole.

In order to observe the combined effects of the adaptation through perceptual semantics with an underlying cross-layer optimization, we observe the individual metrics. The comparison is made between using cross-layer optimization to cope with the network constraints, or no use of it.

Figure 5a shows the value of Δ_α as a function of η and ϵ. In order to achieve high QoS and QoE with the cross-layer optimization, the application layer rate r_{APP} is sacrificed, as more rate is needed to protect from network erasures. Hence, the feasible set of frame rates is reduced, and the obtained α can not achieve the highest expected value. This can be observed with higher values of Δ_α as ϵ increases.

Nevertheless, the cross-layer optimization is guaranteeing very low packet losses, as Figure 5b shows, which translates into minimal artifacts in the video. Hence, while seemingly Δ_α is not as low as expected, the user is guaranteed a seamless video playback.

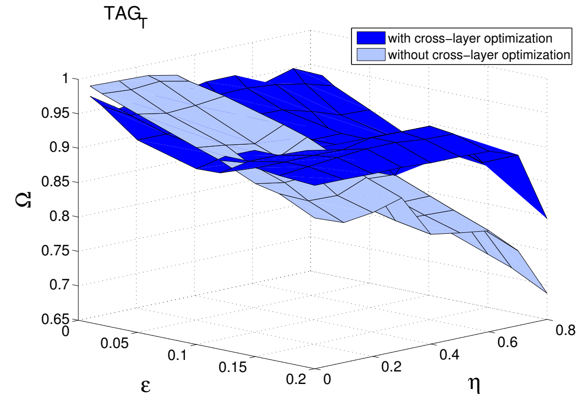


Figure 6. Ω metric for TAG_T

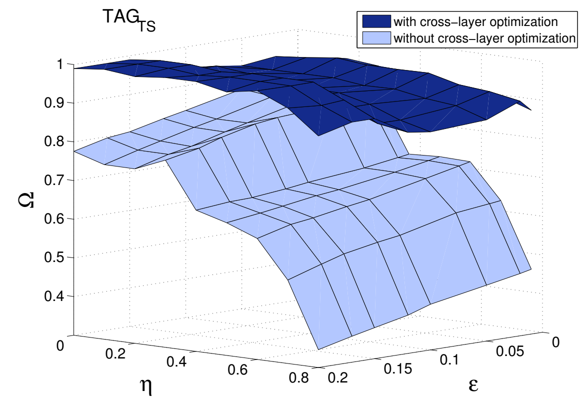


Figure 7. Ω for time-varying semantic tagging TAG_{TS}

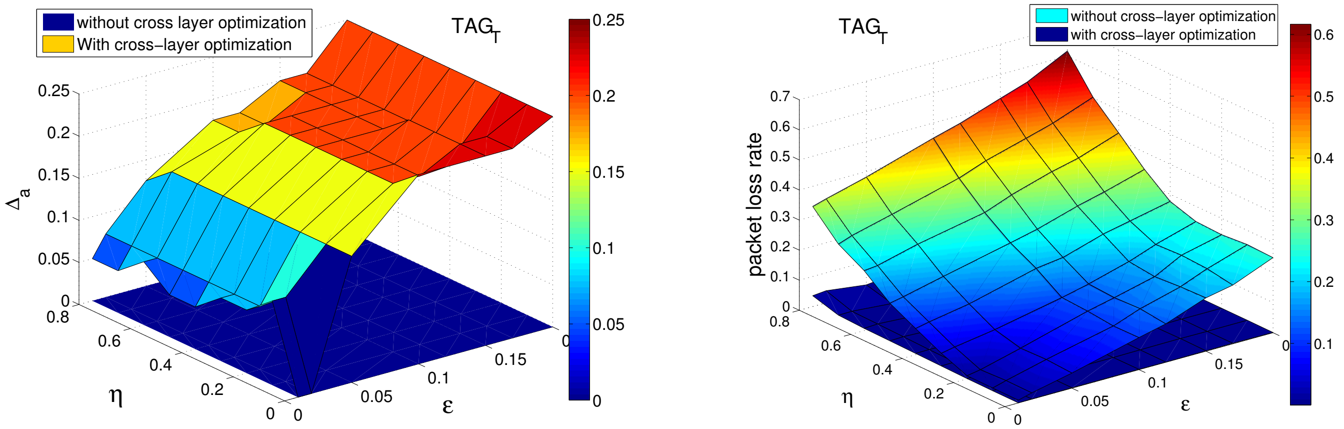
Figure 6 shows the combined metric Ω, where the above trade-off result into higher performance when using cross-layer optimization in combination with the perceptual semantics loop, especially for highly degraded networks.

2) Time varying perceptual semantics tagging: We analyze the effects of time-varying perceptual tagging, representing a realistic case where the user identifies different situations that demand attention towards temporal or spatial features. These variations are represented as alternations of temporal and spatial tagging. Figure 7 shows the performance in terms of the combined metric Ω.

In addition to achieving the expected α demanded through the semantic tagging, the performance is above 80% regardless of the degradations of the network, thanks to the cross-layer optimization. The performance is highly degraded due to congestion, as well as erasures when no cross-layer optimization is used, with performance dropping to 40%. In conclusion, Figure 7 shows that the cross-layer optimization preserves the perceptual semantics.

V. Conclusion and Future Work

We have presented in this paper a framework where we introduced perceptual semantics for video adaptation. Percep-



(a) Δ_α for case 1

(b) $\bar{p}$

Figure 5. Δ_α and $\bar{p}$ for TAG_T

tual semantics are used to acknowledge the user's demand in the context of situation awareness, where special attention is required when using video as means to perceive, comprehend and project ongoing situations, in particular for emergency scenarios. We have presented a novel model for perceptual semantics, based upon these demands, and propose a framework to be integrated into a video adaptive solution, for non-computer aided situation awareness. We discussed how to practically implement perceptual semantics into an adaptive loop that works with an underlying cross-layer optimization in charge of coping with network constraints typical of best effort wireless scenarios. Further, we have shown an adaptive algorithm that translates the perceptual semantics into temporal and spatial resolutions at codec level. Finally, our framework is contextualized for information-centric-networking. Our simulation results show how the perceptual semantic tagging achieves the expected user demands while the underlying cross-layer optimization preserves such performance. Future work includes extensions of perceptual semantics in the ICN context. Moreover, we will study more pertinent QoE metrics to match user's satisfaction when using perceptual semantics. Finally, the presented framework will be further developed for practical usage to implement a potential prototype.

REFERENCES

[1] O. Habachi, Y. Hu, M. van der Schaar, Y. Hayel, and F. Wu, "MOS-based congestion control for conversational services in wireless environments," Selected Areas in Communications, IEEE Journal on, vol. 30, no. 7, 2012, pp. 1225–1236.

[2] M. A. Pimentel-Niño, P. Saxena, and M. A. Vázquez-Castro, "QoE Driven Adaptive Video with Overlapping Network Coding for Best Effort Erasure Satellite Links," in 31st AIAA International Communications Satellite Systems Conference, Florence,Italy, 2013.

[3] H. Seferoglu, A. Markopoulou, U. C. Kozat, M. R. Civanlar, and J. Kempf, "Dynamic FEC algorithms for TFRC flows," Trans. Multi., vol. 12, no. 8, Dec. 2010, pp. 869–885.

[4] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, "Multimedia Streaming via TCP: An Analytic Performance Study," ACM Trans. Multimedia Comput. Commun. Appl., vol. 4, no. 2, May 2008, pp. 1–22.

[5] P. Saxena and M. A. Vázquez-Castro, "Network Coding Advantage over MDS Codes for Multimedia Transmission via Erasure Satellite Channels," in Personal Satellite Services (PSATS), 2013 International Conference on, October 2013, pp. 199–210.

[6] R. Immich, E. Cerqueira, and M. Curado, "Adaptive video-aware fec-based mechanism with unequal error protection scheme," in Proceedings of the 28th Annual ACM Symposium on Applied Computing, ser. SAC '13. New York, NY, USA: ACM, 2013, pp. 981–988.

[7] J. Thomas-Kerr, C. Ritz, and I. Burnett, "Semantic-aware delivery of multimedia," in Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on, Sept 2009, pp. 1498–1503.

[8] C. Henson, A. Sheth, and K. Thirunarayan, "Semantic perception: Converting sensory observations to abstractions," Internet Computing, IEEE, vol. 16, no. 2, 2012, pp. 26–34.

[9] A. Cavallaro and S. Winkler, "Perceptual semantics," in Multimedia Technologies: Concepts, Methodologies, Tools, and Applications, S. M. Rahman, Ed. Information Science Reference, Jun. 2008, ch. 7.5, pp. 1441–1455.

[10] S. B. Kodeswaran and A. Joshi, "Content and context aware networking using semantic tagging," in Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on. IEEE, 2006, pp. 77–77.

[11] J. Strassner, J. Betser, R. Ewart, and F. Belz, "A Semantic Architecture for Enhanced Cyber Situational Awareness," in Secure and Resilient Cyber Architectures Conference, 2010.

[12] M. R. Endsley, "Theoretical underpinnings of situation awareness: A critical review," Situation awareness analysis and measurement, 2000, pp. 3–32.

[13] S. S. Krupenia, C. Aguero, and K. C. Nieuwenhuis, "The value of different media types to support command and control situation awareness," in Proceedings of 9th International ISCRAM Conference, 2012, pp. 1–5.

[14] Z. Ma, F. Fernandes, and Y. Wang, "Analytical rate model for compressed video considering impacts of spatial, temporal and amplitude resolutions," in Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on, July 2013, pp. 1–6.

[15] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications, IETF RFC 3550," United States, 2003.

[16] H. Wirtz and K. Wehrle, "Opening the loops - towards semantic, information-centric networking in the internet of things," in Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2013 10th Annual IEEE Communications Society Conference on, June 2013, pp. 18–24.

[17] C. Tsilopoulos, G. Xylomenos, and G. Polyzos, "Are Information-Centric Networks Video-Ready?" in Packet Video Workshop (PV), 2013 20th International, Dec 2013, pp. 1–8.

On Server and Path Selection Algorithms and Policies in a light Content-Aware Networking Architecture

Eugen Borcoci, Marius Vochin, Mihai Constantinescu,
University POLITEHNICA of Bucharest
Bucharest, Romania
emails: eugen.borcoci@elcom.pub.ro,
marius.vochin@elcom.pub.ro,
mihai.constantinescu@elcom.pub.ro

Jordi Mongay Batalla, Warsaw University of Technology
/ National Institute of Telecommunications
Warsaw, Poland
jordim@interfree.it
Daniel Negru, LaBRI Lab, University of Bordeaux,
Bordeaux, France
daniel.negru@labri.fr

***Abstract* —Appropriate content server and path selection procedures based on different algorithms constitute the first set of actions to be performed in content delivery systems. Multi-criteria optimization algorithms based on user context, network and servers information can be used to enhance the overall efficiency. This paper contains a preliminary work, focused on algorithms and policies for optimized paths and server selection, aiming to finally implement a subsystem in the framework of a content delivery light architecture system.**

Keywords — Content delivery, Server selection, Path selection, Content-Aware Networking, Multi-criteria decision algorithms, Future Internet.

I. INTRODUCTION

The content orientation is an important trend recognized in the current and Future Internet [1] The Information/Content-Centric Networking (ICN/CCN), approach [2][3], revisits some main concepts of the architectural TCP/IP stack. In parallel, “light ICN”. evolutionary solutions introduce Content-Awareness at Network layer (CAN) [4]. Seen partially as an orthogonal solution, Content Delivery Networks (CDNs) improve the content services [5] by distributing the content replica to cache servers located close to groups of users. However, the above solutions involve complex architectures, high CAPEX and significant modifications in Service/Content Providers and Network Providers/Operators.

The DISEDAN Chist-Era project [6][7], (service and user-based **DI**stributed **SE**lection of content streaming source and **D**ual **A**daptatio**N,** 2014-2015) proposes an *evolutionary and light architecture* to enhance the content delivery via Internet. It studies pragmatic solutions for the multi-criteria hard problem of best content source selection, considering user context, servers availability and possibly network status information (if available). The novel concept is based on: a. *two-step server selection mechanism* (at Service Provider (SP) and at End User) by using algorithms that consider context- and content-awareness; b. *dual adaptation mechanism during the sessions*, consisting of media flow adaptation and/or content servers handover. The solution could be rapidly deployed in the market since it does not require complex architecture like ICN, full-CAN or CDN.

This paper contains a preliminary work on paths and server combined selection algorithms and policies applicable by SPs in a light content delivery architecture Section II is a short overview of related work. Section III outlines the overall system and problem description. Section IV contains the main paper contributions, focused on: a. paths and content server selection combined algorithm; b. modifications to allow introduction of SP policies, aiming to increase the system flexibility. Section V contains conclusions and future work outline.

II. RELATED WORK ON MULTI-CRITERIA DECISION ALGORITHMS

This section is a very short overview on some previous work related to *path-server selection* in content delivery systems, based on Multi-Criteria Decision Algorithms (MCDA). The problem belongs to the more general one known as *multi-objective optimization*. This has been extensively studied in various and large contexts of economics and engineering. The paper will not detail this. Few references are given at the end of the paper [8][9][12].

The general problem of multi-objective optimization is to find *min F(x)* = $[f_1(x), ..f_k(x)]$ where $x \in X^i$, the decision variables space, and $f_1(x), ..f_k(x)$, are a set of objectives, [8] [9]. Such problems are in general NP complete, so, different simplified heuristics have been searched. A simple scalar approach maps the k-dimensional vector onto a single scalar value *w* by using an appropriate cost function c(), thus reducing the problem to a single-criterion one. However, information about individual components is lost. In the server-path selection problem, several decision parameters are important, such as: server load and proximity, transport path (length, bandwidth, loss, and jitter).

Solutions have been searched treating the decision variables separately and considering them as independent. Note that in our case this is only partially true, e.g., delay and jitter are clearly not independent variables. Therefore modifications should be added to the basic algorithm to capture such effects and this paper proposes a solution.

The *reference level decision algorithms*, [10][11], considers a decision space R^m and the decision parameter/variables: v_i, i=1, ..m; $\forall i$, $v_i \geq 0$. A candidate solution is an element $S_s=(v_{s1}, v_{s2}, .., v_{sm}) \in R^m$. Let S be the number of candidates indexed by s = 1, 2, ..S. The value ranges of decision variables might be bounded by given

ISBN: 978-1-61208-368-1

constrains. The selection process searches a solution satisfying a given objective function, conforming a particular metric.

The basic algorithm defines two reference parameters:

- r_i =*reservation level*=the upper limit for a decision variable which should not be crossed by the selected solution;
- a_i=*aspiration level*=the lower bound for a decision variable, beyond which the solutions are seen as similar.

Without loss of generality one may apply the definitions of [11], where for each decision variable v_i there are defined r_i *and* a_i , by computing among all solutions s = 1, 2, ..S:

$$r_i = \max [v_{is}],\ s = 1, 2, ..S$$
$$a_i = \min [v_{is}],\ s = 1, 2, ..S \qquad (1)$$

In [11], modifications of the decision variables are proposed: *replace each variable* with *distance from it to the reservation level*: $v_i \rightarrow r_i\text{-}v_i$; (increasing v_i will decrease the distance); normalization is also introduced to get non-dimensional values, which can be numerically compared. For each variable v_{si}, a ratio is computed, for each solution *s*, and each variable i:

$$v_{si}' = (r_i\text{-}v_{si})/(r_i\text{-}a_i) \qquad (2)$$

The factor $1/(r_i\text{-}a_i)$ - plays also the role of a weight. The variable having high dispersion of values (max – min) will have lower weights, and so, greater chances to determine the minimum in the next relation (3). In other words, less preference is given to those variables having close values (reason: selection among them will not influence significantly the overall optimum). The algorithm steps are:

Step 0. Compute the matrix M$\{v_{si}'\}$, s=1…S, i=1…m

Step 1. Compute for each candidate solution *s*, the minimum among all its normalized variables v_{si}':

$$min_s = min\{v_{si}'\};\ i=1...m \qquad (3)$$

Step 2. Make selection among solutions by computing:

$$v_{opt} = max\,\{min_s\},\ s=1, ..S \qquad (4)$$

This v_{opt} is the optimum solution, i.e it selects the best value among those produced by the Step 1.

The reference level algorithm has been used in several studies.

The work [13] proposes a decision process for network-aware applications, based on reference level MCDA with several metrics. The improvement (compared to the basic algorithm) consists in considering not only the currently selected server status, but also the system future state after the selection. The simulation results showed a slight gain versus the basic algorithm, while using the same information from the network level (server and link load).

The work [14] proposes and evaluates a multi-criteria decision algorithm for efficient content delivery applicable to CDN and/or ICN. It computes the *best available source and path* based on information on content transfer requirements, servers and users location, servers load, and available paths. It runs processes at two levels: 1. *offline* discovering multiple paths, and gathering their transfer characteristics; 2. for each content (online) request, finding the best combined server – path (reference level model). The following "use cases" are evaluated: *random server and random path*, combined with shortest single path routing protocol (current Internet solution); *closest server and random path*, (similar to the current CDN); *least loaded server and random path*; *best server and the path with more available bandwidth* in the bottleneck link. Simulation, using Internet large scale network model, confirmed the effectiveness gain of a content network architectures (i.e., having a degree of network awareness) and efficiency of the combined path-server selection.

The work [15] models and analyzes a simple paradigm for *client-side server selection*. Each user independently measures the performance of a set of candidate servers, randomly chooses two or more candidate and selects the server providing the best hit-rate. The algorithm converges quickly to an optimal state where all users receive the best hit-rate (respectively bit rate), with high probability. It is also shown that if each user chooses just one random server instead of two, some users receive a hit-rate (respectively, bit rate) that tends to zero. Simulations have evaluated the performance with varying choices of parameters, system load, and content popularity.

The contributions of this paper w.r.t. previous work mentioned are summarized as: two-phase flexible selection procedure based on MCDA reference level algorithm, applicable with slight modifications for nine use cases (see Section IV); additional policy supporting modifications proposed for the basic algorithm, in order to capture different Service Provider strategies.

III. DISEDAN SYSTEM SUMMARY

The DISEDAN solution performs an initial path-server selection and then, during the session, applies media flow adaptation based Dynamic Adaptive Streaming over HTTP (DASH) and/or content server handover. Details are described in [6], [7]. The system has a light architecture in the sense that it does not mandatory assume special Management and Control Plane at SP and end user sides. However the SP can provide to the client, at least a list of available and appropriate servers and/or other (offline and/or online observed) information to optimize the final selection at EU side selection results The design is backwards-compatible: both (un)modified client and/or SP can cooperate. Based on the evaluated current delivery conditions, rules are defined to decide which adaptation action to perform. The DISEDAN implementation will be flexible [6], [7], allowing cheap and seamless deployment.

This paper is focused on the path-server selection problem, applicable to DISEDAN. The acquisition of the input information for the selection procedure is out of scope of this work; it is supposed that such information is provided statically or dynamically (by measurements) and made available for the algorithm.

IV. PATH AND SERVER SELECTION OPTIMIZATION

A two phase selection process is adopted here, similar to [14]. The Phase 1 is executed offline and computes candidate paths from servers to users. The Phase 2 applies a MCDA (reference level variant) algorithm and computes the best path-sever solution, based on multi-criteria and also policies guidelines. Note that the multicriteria algorithm is flexible: any number of decision variables can be used, depending on their availability. For instance in a multi-domain network environment it is possible that SP has not relevant or complete knowledge about end to end (E2E) transport paths. In such cases the list of available decision variables can be as well used. Another additional contribution here consists in modifying the reference algorithm, to include different SP policies concerning the importance of some decision variables with respect to others.

i. Network Environment

The content delivery for large communities of users frequently involve several network domains independently managed, [4][5]. In a combined optimization procedure for path and server selection it is not realistic, from the real systems management point of view, to consider all details of the paths from the content servers to the users. Therefore (supposed in this paper and also in DISEDAN), the network awareness of the management and control entity of an SP is limited, e.g. to knowledge about the inter-domain context, i.e., the inter-domain graph (where each network domain is abstracted as a node) and inter-domain link capacities, while considering the multi-tier organized Internet. The location (domains) of the potential groups of users and server clusters are also supposed to be known.

Figure 1 shows a generic example of a tiered structure network, containing several domains D11, ..D33 interconnected via inter-domain links. At the edges of this structure, groups of servers and users are connected to Tier 3 domains. In Figure 1, two possible paths from D33 to D32 are shown. The Phase 1 procedure will compute such similar paths between two edge domains.

ii. Use cases for path-server selection procedure based on MCDA algorithm

Several Use Cases can be defined for a combined algorithm, by considering several criteria for the path and server selection. Several metrics can be defined for paths and servers status evaluation. The path metric can be the simplest - number of hops, or a more powerful one (enabling better QoS assurance), e.g.: link cost=1/B, where B could be the static link capacity or the available bandwidth (dynamically measured). Also constrained routing policies can be applied (e.g. related to bandwidth, number of hops, etc.).

The bandwidth of the selected link should be the maximum one (among several paths) but evaluated at the bottleneck link of that path. Additionally, the path might be constrained, e.g.: the number of hops (i.e., domains), should be lower than a maximum. For server status, one could consider the server proximity to the user, or server load. The MCDA algorithm has the quality that it can use several decision variables and make a global optimization.

For the path selection one may apply: a. *Single path between server and user* (usually provided by the current Internet routing based on minimum number of hops); b. *Random path selected among equal costs paths* between server and user, given that a multipath protocol is applied (e.g modified Dijkstra algorithm); c. *best path among several paths* having similar costs in a defined range.

For server selection one may apply: 1. *Random selection*; 2. *Closest server* to the user (e.g., considering as metric the number of hops i.e domains - between server and user; 3. *Least loaded server* (the load can be evaluated as the current number of connections, or partially equivalent- as the total bandwidth consumed at the server output).

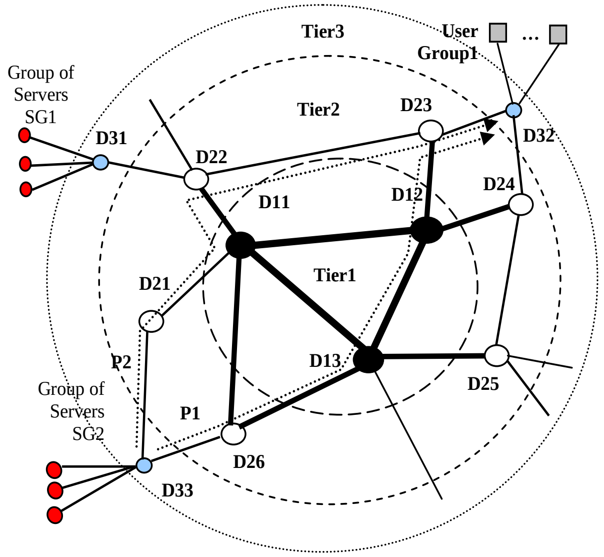


Figure 1. Example of a sample tiered network. P1 and P2 – paths from domain D33 to D32.

Considering combinations of the above factors, nine Use cases (and corresponding algorithms) can be defined: a.1, a.2, ...c.3, if independent decisions are taken for path, and respectively the server, with no MCDA algorithm. However we will consider a global optimization MCDA algorithm with several decision variables taken from the above.

iii. Two phases path-server selection procedure

The following simplifying assumptions are considered valid for this first version of the selection procedure:

- All servers are managed by the unique Resource Allocator (RA) belonging to SP Manager. The RA knows each server status, including its current load (number of active connections and bandwidth consumed at the server output). A degree of content-awareness exists in RA; it knows the inter-domain graph, and inter-domain link capacities.
- Each domain is considered as a node in the network graph, i.e. the intra-domain transport is not visible. This is a major realistic assumption in simplifying the amount of knowledge supposed to exist at SP level.
- All servers and users location are established offline, and are fixed. However the system can accommodate

ISBN: 978-1-61208-368-1

the end user terminal mobility, given that in the content delivery phase a content server switching is possible.

- The total number of content objects (*Max_no_CO)* are distributed (offline mode, by an external caching process, out of scope of this algorithm) to server groups and between the servers of a given group, while the number of COs in a server should be ≤ *Max_no_CO_per_Server*.
- The content object instances replicated in surrogate servers are known by the RA. A data structure *CO_SRV _map* contains the mapping of CO replica on servers. Each CO is stored in 1, 2…. K servers; K = maximum number of servers to replicate a content object.
- The time-life of a CO instance in a server is unlimited.
- All COs are delivered in unicast mode, so a “connection” is 1-to-1 mapped to a content consumption session. The COs have the same popularity.
- Each CO user request asks for a single CO; however the same CO can be consumed simultaneously by several users, by using private connections.
- RA treats the User requests in FIFO (queue named *COreq_Q*) order.
- RA accepts or rejects user requests. Rejection happens if there are no servers, or no transport resources available. No further negotiation between the User and RA is assumed after a request transaction processing.
- The bandwidth occupied by a connection is equal to *Bw_CO* (in the first approach it can be considered constant). More generally this bandwidth is random, in a range Bw_CO +/- ΔBw.
- A connection load for the server and path will be Bw_CO, during Tcon interval measured from the connection request arrival instant (we neglect the processing time for content/connection requests).
- RA uses the most simple additive bandwidth management (no statistical multiplexing is assumed).
- The average duration of a connection (for content consuming) is Tcon. The real duration could be in a range TCon +/- ΔTcon.

Phase 1

The Phase 1 (offline) general objective is to compute, on the inter-domain graph, (multiple) paths from server domains to user domains. No traffic load consideration is applied. The input data are: topology, inter-domain link capacities, location of servers, and users. Some constraints can be applied, e.g., bottleneck bandwidth (BB) on any path ≥ Bmin; number of hops (domains) on any path ≤ NHmax. The simplest metric is the classic one (number of hops). More powerful approaches compute multiple paths: equal cost paths, or sets of paths having costs in a given range. Having more than one path would provide several MCDA choices opportunity. The multiple paths can be computed, by running a modified version of the classic Djikstra algorithm [16]. A “better” (from QoS point of view) additive metric is: *link_cost= 1/*B_{link}, where B_{link} is the link bandwidth/capacity). Given that routing process is a classic one, it will be not detailed in this paper. The Phase 1 output is a set of sub-graphs, each one containing the multi-paths from a given group of servers to a given group of users. The Phase 1 algorithm is convergent. Its order of complexity is not higher than for different variants of Dijkstra based algorithms, [17].

Phase 2

The Phase 2 of decision process jointly selects (for each user request arrived at RA), the best server and path (based on dynamic conditions) from the available candidates computed in the Phase 1. The signalling details user-RA are out of scope of this paper. The RA applies an admission control decision, followed/combined with an MCDA algorithm. The Phase 2 dynamicity means updating the paths and server loads according to the new requests arrived. Also considering the time-life of a connection, different server status items are updated when the connections are terminated. Note that there is no problem to downgrade the algorithm if complete path information is missing. More generally, the number of decision variables and the amount of information existent on them (static and/or dynamic) are flexible items.

A description of Phase 2 is given below.

Request analysis and resource allocation (pseudocode)

// It is assumed a time process which triggers activation of the main procedure, at each generic time tick instant Tk. This approach can serve also for managing the time lives of connections. The algorithm description is given below.

```
    Each Tk
    {   While COreq_Q non-empty
            {req = Extract_first_element_from COreq_Q( );
             Process_request (req);//processes the first request from the COreq_Q
            Adjust_time_life_of_connections_in servers; }
     }
Process_request(req) // description of a user request processing
        {Identify_Server _groups_and_individual_servers_able_to_provide_CO ; // candidate servers for requested content
                {//Search in the CO_SRV _map, by using the CO index in the request};
        Create _candidate_servers_vector; // containing one entry for each such server
        Collect_status_of_each_server;  //from a data structure Server_status, the status of each sever is loaded in the
                                        // vector; in the most simple variant : the current number of active connections
        Determines_ sub-list_of _paths_for each candidate_server; // from the list of updated paths, by using information
```

```
                                                  // from the Phase 1
          Create_candidate_list_of_path_server_solutions;//each solution is characterized by server load, bandwidth
                                                  // and number of hops
          Delete_full_loaded_servers; //optional; it can be included in MCDA algorithm
          Delete_elements from_the_list_of_paths_associated_to_the_candidate_list:// optional; it can be done by MCDA
                // those which have number of hops > NHopmax;
                // those which have Available Bandwidth < Bmin;
          Run_the_MCDA reference_level_ algorithm ;// determine best path-server solution; policies can be included here
          If successful
             then
               {Increase_success_list_statistics;
                Update_the_allocated_server_load;
                   // Increase the number of active connections
                   Load & start timer associated to the time-life of this connection
                Add_additional_bandwidth_consumed_to_ the_allocated_path_load on all links;}
             else increase the reject list statistics;
          }
Adjust_time_life_of_connections // delete the terminated connections from the server status
          For each server //Sv1, …Svn
            { For each timer
               { If Active_flag=1 and Timer_value >0
                   then {Timer_value - -;
                        If Timer_value = 0 then {Active_flag=0; NCO_srv --;}}}
Generation_Content_object_request_
          Initialization: TReq = random [1,…P*Tk];
          Each Tk // equivalent with periodic interrupts at Tk seconds interval
                {Treq = Treq - 1;
                   If Treq =0   then
                      { k = random [1, …. Max_no_CO];
                      Put_CO_req (User_id, Tcon, COk,)_in_COreq_Q;//generate content object request
                      TReq = random [1,…P*Tk]}; // restart timer and select a random interval until the
                                                  // next request generation}
```

Policy guiding the MCDA

Several remarks can be done related to the basic reference level algorithm:

- The formula $min_s=min\{ v_{si}' \}; i= 1..m$ *(3),* selects as *representative of each candidate solution*, the "worst case" value, i.e., for all other variables/parameters, this solution has "better" normalized values then this representative. This is arithmetically correct, however in practice this "worst" case parameter might be actually less important than others, either from technical or business (i.e policies) point of view.
- In some particular cases with dependent variables (e.g., delay/jitter) the solution selected could be not the most appropriate, from actual implementation point of view.
- The step 2 *compares values coming from different types of parameters* (e.g., 1/Bwdth, delay, jitter, server load, etc.) - independent or dependent on each other. The normalization allows them to be compared in the *max{ }* formula. However, the numbers compared are from items having different nature. *This is an inherent weak property of the basic algorithm.*
- More important is that the SP might want to apply some policies when selecting the path-server pair for a given user. Some decision variables could be more important than others. For instance the number of crossed domains (no_of_hops in MCDA) can be the most important parameter – given the transit cost. In other cases the server load could be more important, etc.

A simple modification of the algorithm can support a variety of SP policies. We propose here a modified formula:

$$v_{si}' = w_i(r_i-v_{si})/(r_i-a_i) \qquad (3’)$$

where the factor $w_i \in (0,1]$ represents a weight (priority) that can be established from SP policy considerations, and can significantly influence the final path-server selection. This will solve the above mentioned issues.

A sample example below shows the optimization obtained. Let us consider a selection scenario in which the decision variables are given in Table 1, and six candidates in Table 2 (entries are native not-yet normalized values)

Priorities are introduced in Table 1, derived from SP policy. Here, the server load and numbers of hops are considered the most important.

One can define: a1= 0, r1=100; a2=0, r2=10; a3=110, r3 = 10; a4=0, r4=50; a5=0, r5=100.

TABLE I. DECISION VARIABLES EXAMPLE

Decision variables	**Semantics**	**Units**	**Priority**
v1	server load	(%)	1- max
v2	number of hops	Integer	1
v3	available bwdth on the path	Mbps	2
v4	jitter	ms	3
v5	E2E delay	ms	4- min

TABLE II. CANDIDATE SOLUTIONS EXAMPLE

	s_1	s_2	s_3	s_4	s_5	s_6
$\boldsymbol{v_{s1}}$	0	20	40	70	80	100
$\boldsymbol{v_{s2}}$	5	7	6	3	4	5
$\boldsymbol{v_{s3}}$	40	20	50	80	50	60
$\boldsymbol{v_{s4}}$	0	10	30	20	10	30
$\boldsymbol{v_{s5}}$	30	80	70	40	30	50

Applying the basic algorithm (i.e., with no priorities) simple computation will show that formula (4) is *max{0.3, 0.1, 0.3, 0.3, 0.2, 0},* showing that solutions s1, s3, s4 are equivalent. However, examining the initial input candidate values, it is clear that s_1 is the best (server load=0, and sufficient available bandwidth- compared to others).

Now, we introduce policies, assuming the priorities assigned in Table 1. Some weights (acting as compression factors) can be defined, e.g., $w_1= 0.5, w_2= 0.5, w_3= 0.7, w_4= 0.8, w_5= 1.0.$ Then applying the formula (3'), one gets a new set of values for the formula in (4), i.e., *max {0.21, 0.07, 0.2, 0.15, 0.1, 0}.* It is seen that s1 solution is now selected as the best, which corresponds to the intuitive selection of it.

Some other examples have been checked to verify the prioritized selection capability of the modified MCDA. Note that despite its simplicity the modification proposed can have major impact on algorithm results, given that different SP policies can be defined, depending on user categories, content server exploitation needs, networking environment, etc. Therefore, the weighting factors in practice do not come from some formulas, but should be chosen, based on the defined priorities of the SP. A natural usage of the modified algorithm proposed here could be to select several sets of best solutions, fit to the different policies of the Service Provider.

V. CONCLUSIONS AND FUTURE WORK

This paper presented a preliminary study on multi-criteria decision algorithms and procedures for best path-server selection in a content delivery. While applying some previous ideas of two phases procedure (offline and online) the solution adopted here is a flexible (supporting many use cases) modified decision procedure which additionally can capture some policy related priorities for decision variables. It was shown that such modifications can enhance the added value of the decision taken by the algorithm.

Future work will be done (in the DISEDAN project effort) to simulate the system in a large network environment, and finally, to implement the described procedures in the framework of a system dedicated to content delivery based on a light architecture.

ACKNOWLEDGMENTS

This work has been partially supported by the Research Project DISEDAN, No.3-CHIST-ERA C3N, 2014- 2015 and project POSDRU/107/1.5/S/76903.

REFERENCES

[1] J. Pan, S. Paul, and R. Jain, "A survey of the research on future internet architectures", IEEE Communications Magazine, vol. 49, no. 7, July 2011, pp. 26-36.

[2] J. Choi, J. Han, E. Cho, T. Kwon, and Y.Choi, "A Survey on Content-Oriented Networking for Efficient Content Delivery", IEEE Communications Magazine, March 2011, pp. 121-127.

[3] V. Jacobson, et al., "Networking Named Content," CoNEXT '09, New York, NY, 2009, pp. 1–12.

[4] FP7 ICT project, "MediA Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments", ALICANTE, No. 248652, http://www.ict-alicante.eu, Sept. 2013.

[5] P. A. Khan and B. Rajkumar. "A Taxonomy and Survey of Content Delivery Networks". Department of Computer Science and Software Engineering, University of Melbourne. Australia : s.n., 2008. www.cloudbus.org/reports/CDN-Taxonomy.pdf.

[6] http://wp2.tele.pw.edu.pl/disedan/, retrieved: 07, 2014.

[7] http://www.chistera.eu/sites/chistera.eu/files/DISEDAN%20-%202014.pdf, retrieved: 07, 2014

[8] R. T. Marler, and J. S. Arora, "Survey of multi-objective optimization methods for engineering". Struct Multidisc Optim Eds., No. 26, 2004, pp. 369–395.

[9] J. Figueira, S. Greco, and M. Ehrgott, "Multiple Criteria Decision Analysis: state of the art surveys", Kluwer Academic Publishers, 2005

[10] J. R. Figueira, A. Liefooghe, E-G. Talbi, and A. P. Wierzbicki "A Parallel Multiple Reference Point Approach for Multi-objective Optimization", "European Journal of Operational Research 205, 2 (2010), pp. 390-400", DOI: 10.1016/j.ejor.2009.12.027.

[11] A. P. Wierzbicki, "The use of reference objectives in multiobjective optimization". Lecture Notes in Economics and Mathematical Systems, vol. 177. Springer-Verlag, pp. 468–486.

[12] T. Kreglewski, J.Granat, and A. Wierzbicki, "A Dynamic Interactive Decision Analysis and Support System for Multicriteria Analysis of Nonlinear Models", CP-91-010, IIASA, Laxenburg, Austria, 1991, pp 378-381.

[13] J. M. Batalla, A. Bęben , and Y. Chen, "Optimization of the decision process in Network and Server-aware algorithms", NETWORKS 2012, October 15–18 2012, Rome.

[14] A. Beben, J. M. Batalla, W. Chai, and J. Sliwinski, "Multi-criteria decision algorithms for efficient content delivery in content networks", Annals of Telecommunications - annales des telecommunications, vol. 68, Issue 3, 2013, pp. 153-165, Springer.

[15] C. Liuy, R. K. Sitaramanyz, and D. Towsleyy, "Go-With-The-Winner: Client-Side Server Selection for Content Delivery", http://arxiv.org/abs/1401.0209, retrieved: 07, 2014

[16] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, "Introduction to Algorithms", The MIT Press, Cambridge, Massachusetts, 2000, ISBN: 0-262-53091-0.

Building Efficient End-to-End Service Transparent Fiber Networks Supporting Access Rates Beyond 10Gb/s

Theofanis G. Orphanoudakis, Chris Matrakidis,
Christina (Tanya) Politi, Alexandros Stavdas
Dept. of Informatics and Telecommunications
University of Peloponnese
Tripolis, Greece
{fanis, cmatraki, tpoliti, astavdas}@uop.gr

Helen-Catherine Leligou, Evangelos Kosmatos
Dept. of Electrical Engineering
Technological Educational Institute of Central Greece
Psahna—Evia, Greece
leligou@teihal.gr

Abstract— **Currently, networks are developed based on the layered model focusing on customized solutions for the access, metro and core domains. Access networks rely upon centralized packet multiplexers, while core network switches rely on large port count electronic switches assisted where appropriate by Optical Cross-Connects (OXCs). However, this monolithic model can lead to capacity and space bottlenecks, compromise network services because of a lack of common feature sets, and limit the revenue that data services can produce for operators. In this paper, we present a novel control plane solution, which can lead to metro network segment collapse and access-core integration supported by an efficient traffic control and distributed multiplexing scheme based on a hybrid long-reach fiber access network architecture. We show that this architecture can be exploited as a large scale distributed multiplexer that can be used to funnel traffic directly from access networks over a core optical network and describe a control plane architecture compatible with the concept of Software Defined Networking for simplifying the aggregation process and improving performance at the same time.**

Keywords-Passive optical networks; wavelength division multiplexing; access-core integration; medium access control; FTTx.

I. Introduction

Telecommunications traffic is soaring at an astonishing rate. Worldwide, the average traffic will increase threefold over the next five years [15]. The evidence we have so far indicates that the demand for even higher capacity networks is steadily increasing. The widespread availability of bandwidth intensive services and advanced Fiber to the x point (FTTx) schemes will have dramatic consequences in core networks: In this paper, we propose a network control and data plane architecture aiming to capitalize on the complementary strengths of “optical” and “electronics” technologies so as to design an ultra high capacity end-to-end network solution allowing for transparent core-access integration. The existing and new services/applications that are becoming available to the end user are currently supported by wireline access technologies, such as relatively limited capacity cable modems, and Digital Subscriber Line (DSL) [14] or high capacity FTTx/Very-high-bit-rate DSL (VDSL) [13] as well as wireless technologies such as mobile 3G/4G/Long Term Evolution Advanced (LTE-A) networks (High-Speed Downlink Packet Access (HSDPA)/High Speed Packet Access (HSPA)+) [12], WiFi and WiMax. Wireless networks feature high flexibility in terms of broad area coverage but face the limitations of relatively low bandwidth (10’s to a few 100’s of Mbps shared between all the users in a cell). Therefore, next generation fiber access networks are ultimately expected to become a universal access networking platform for broadband service delivery either directly to end-users or as a wireless/mobile access backhaul infrastructure.

While deployment of dedicated fibers per subscriber may prove economically unjustifiable, there are several solutions that may lead to resource sharing, hence, cost reduction. Passive Optical Networks (PONs) were initially proposed in 1980s to efficiently concentrate/distribute traffic via a commonly shared, passive tree-shaped topology. Under this scheme, the time sharing of this topology allowed traffic from multiple Optical Network Units (ONUs) to reach, collision free, a single port at the Optical Line Termination (OLT) [1]-[3].

In the core network, L2/L3 switches and routers are progressively pushed back to core network periphery. Thanks to architectural changes in access networks a trend for significant core node consolidation is emerging i.e., fewer, but higher capacity, nodes across the network. Hence, efficient ways to aggregate and transport traffic over the core network infrastructure are required. The objective is to exploit the immense capacity offered by optical transportation systems, while avoiding the cost of electronic switching at transit nodes. This task in turn requires appropriate traffic grooming and forwarding schemes that can operate at the optical layer and reduce conversions to the electronic layer as much as possible to reduce costs. In order to achieve this, a coordinated end-to-end network operation is required. This coordination must take into account the boundaries of different administrative domains and provide appropriate interfaces to allow exchange of information and implementation of a distributed information forwarding and traffic aggregation scheme.

In the rest of this paper, we describe the framework for developing such a network architecture. In the following section, we focus on the end-to-end network view and the core network functionality. In Section III, we describe an interoperable long reach access network architecture that can lead to access-core integration. In Section IV, we describe an integrated resource reservation scheme that can operate over both Wavelength-Division Multiplexing (WDM) and Time-Division Multiplexing (TDM) shared optical networks. In Section V, we evaluate the proposed scheme under specific scenarios. Finally, Section VI concludes our paper.

II. End-to-End Network Architecture

The objective of an end-to-end network is to collect the traffic from the access part and forward it to the recipient access network while providing the requested Quality of Service (QoS) performance. Our approach is to optimize the performance in this segment proposing suitable network and node architectures in the framework of dynamic networking while ensuring backwards compatibility. The overall vision is shown in Figure 1. The main ideas are to remove the need for a physical aggregation network, use optical interconnection between access and core whenever possible, and still reap the benefits of statistical multiplexing to achieve efficient use of resources, i.e., exploiting access-core integration [4].

In order to achieve the objective of efficiency (i.e., reduced deployment and operational costs), the network must implement efficient traffic aggregation and routing schemes so that the available fibers and switch ports are utilized to the greatest possible extent. Additionally, the transparency at the optical layer should be maintained to the greatest possible extent as well, so as to reduce the increased cost of deploying complex and costly in terms of power consumption packet routers. Transparency could be achieved in different forms. Ultimately, optical transparency could guarantee traffic forwarding directly at the optical layer. In this case scalable optical switches could be exploited. However, this is difficult to achieve since traffic flows need to be redirected based on flexible rules that lead to specific requirements about traffic processing at each node. While optical transparency is difficult to achieve end-to-end in long paths, we will show that the same objectives can be achieved if service transparency is maintained. By service transparency, we refer to aggregation and processing of traffic flows aggregated and managed in terms of provisioned network services that can lead to reduced complexity and efficient implementations. Thus, core optical nodes can be exploited for transportation and switching of large traffic containers at extremely high data rates. Network and node architectures to achieve these objectives has been described in [5][6].

This could exploit a flexible and scalable control plane to allow exchange of information across domains so as to achieve implementation of optimal traffic aggregation rules so that well utilized optical flows are switched end-to-end across network segments remaining entirely in the optical domain. In the example of Figure 1, under appropriate coordination of reservations in the access domain (1) of the source node traffic from different user interfaces and services can be aggregated and transported over appropriately formed L2 traffic containers. Such containers can then reserve wavelength resources (possibly time-shared) over the core domain (2, 3 in Figure 1) and reach the recipient over appropriately reserved resources at the destination access domain (1' in Figure 1).

To achieve optimal resource utilization in the network,

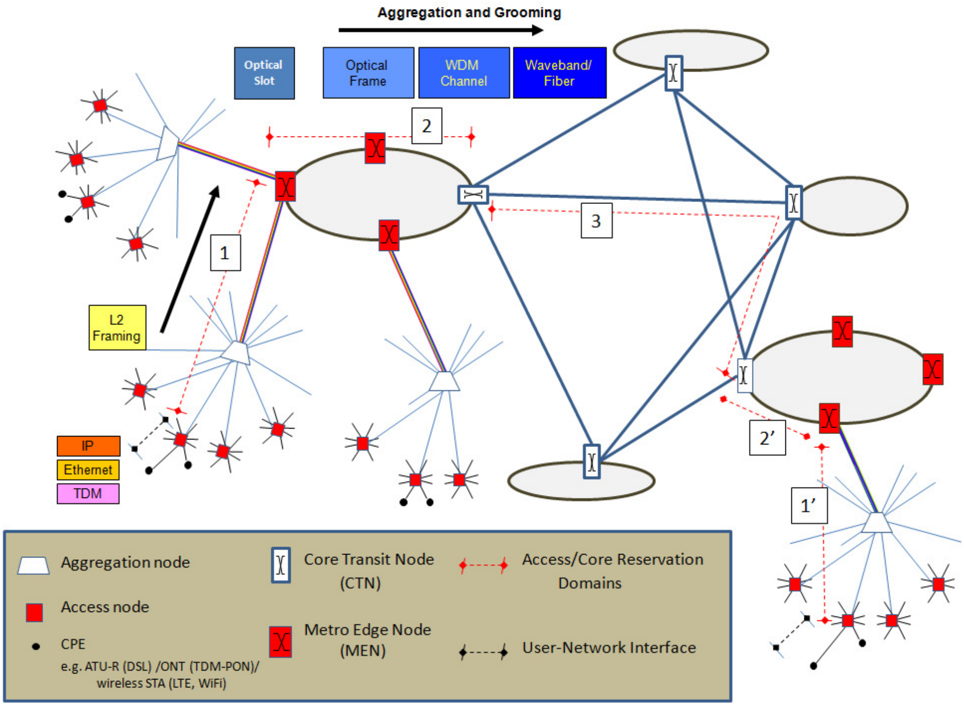


Figure 1. End-to-end network architecture and resource reservation domains.

however, the end-to-end paths signaled by the network control plane must be computed based on available network status information at each segment. Taking into account the requirement for interoperability across administrative domains this can be addressed through the introduction of an appropriate path computation architecture, where peering or hierarchical Path Computation Elements (PCEs) will cooperate for the computations of end-to-end optimal paths, even in scenarios spanning different administrative domains (Figure 2). Such a distribution of the path computation functionality network serves a two-fold purpose. First, by deploying a PCE per network segment (e.g., one per cluster and another in the core part connecting CTNs), a scalable path computation tailored to the characteristics of each network segment can be implemented. Furthermore, assuming that network segments are owned by different network operators, optimal end-to-end path computations without compromising the confidentiality between domains can still be achieved. For instance, this would likely be achieved through the deployment of hierarchical PCE architectures, or the usage of path-key mechanisms amongst peering PCEs.

The above scheme suggests a seamless integration of abstraction and resource orchestration mechanisms across the entire physical layer infrastructure. Towards this end, new technologies for optimal configuration and planning to reduce costs, simplify management, improve service provisioning time, and improve resource utilization in multi-domain networks become increasingly important. Ideally, an application or service could be completely decoupled from the underlying network infrastructure, but this is not always realistic. In most cases, access to the specific underlying infrastructure quality performance indicator factors and resource management mechanisms, is required in order to establish and manage specific Service Level Agreements (SLAs). To meet application performance objectives, it becomes necessary for the application or its proxy to ensure that the underlying network is aware of the application requirements and provides the necessary services. This is a task of the network control and management plane entities in order to appropriately map applications to offered services configure the network elements and orchestrate reservation of resources across network segments. While proprietary implementations exist to achieve the interactions described above most are based on some sort of a network control plane that utilizes specific interfaces and protocols. The objective is to decouple network processing and information forwarding data plane functions from the service composition, configuration and management functions. To support transport networks with multiple administrative and technology segments, communication and interoperability between control planes is required. The mediation between control and transportation, layers can facilitate joint optimization of computing and networking deployments, software-defined functions at a number of layers, including the transport (i.e., optical and TDM) layer, enabling simpler interworking of different administrative and technology domains, and application-aware transport network resources.

Motivated by the above objectives and in order to replace proprietary implementations and manual processes with an automated control plane mechanism the Software Defined Networking (SDN) paradigm has emerged [7][8]. The recently introduced SDN concept based on the OpenFlow [8] protocol relies on a complex notion of a flow associated to a number of header fields in the data frame, which enables the definition of detailed rules for the treatment of different traffic classes and owners. Because of this, SDN is seen as a potential solution for a unified control plane in converged access/aggregation and mobile networks. Thus,

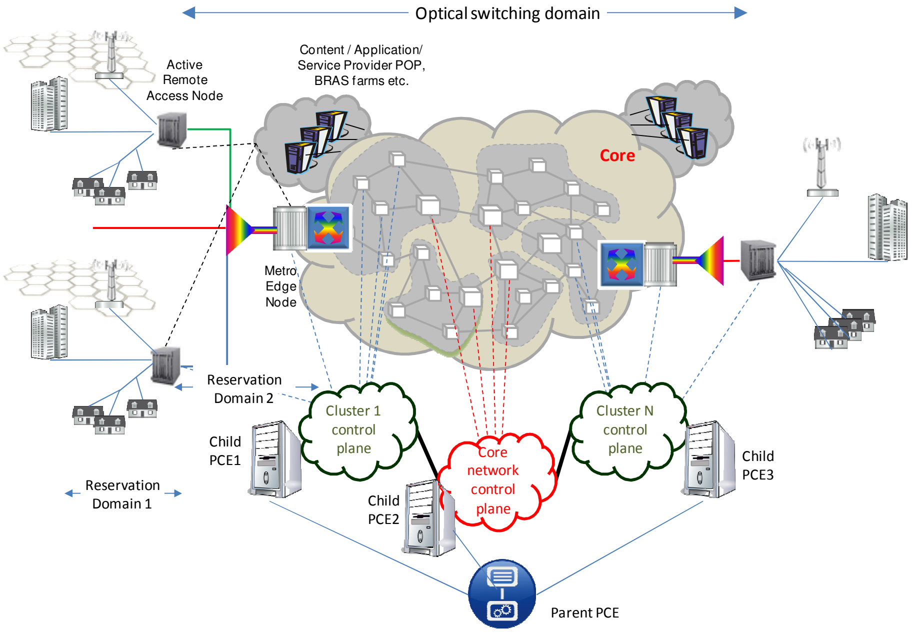


Figure 2. SDN based control plane, hierarchical scenario.

SDN principles could be employed to implement service delivery policies and coordinate end-to-end reservations implementing the framework described above resulting in efficient service transparent optical networks.

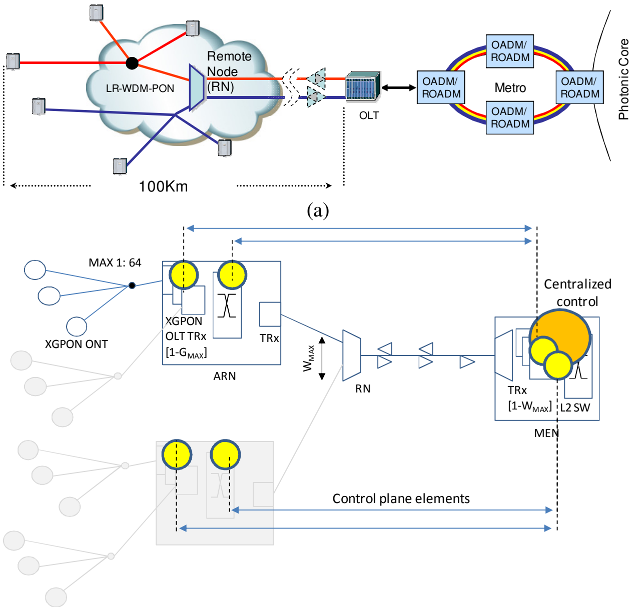
III. ACCESS NETWORK ARCHITECTURE

Having described the vision for achieving service transparency over the core network, we should also focus on how to achieve deeper fiber penetration in the last mile building an end-to-end optical infrastructure maintaining service transparency so as to increase efficiency and reduce complexity. Towards this end, we first describe the available technologies and architectural components that can be used to develop dynamic optical access networks and present a scalable reference access network architecture. In the following section, we will describe how the control plane mechanisms described above can be extended to support end-to-end resource reservation schemes across the entire optical network.

WDM-PONs, possibly complemented by TDMA techniques, are considered the next step in the evolution of PONs. They can lead to higher per-ONU bandwidths, splitting ratios, and reach, as compared to EPON and GPON architectures due to the higher capacity per fiber. The use of WDM-PONs enables new broadband business and residential applications on a broad scale, and enables the evolution of metro area networks towards a unified access and backhaul infrastructure. Different per-wavelength bit rates ranging from 1 to 10 Gb/s have to be supported, and full integration into a management system and also into a control plane is necessary.

A WDM-PON architecture mainly depends on the use of the so-called Remote Nodes (RNs), which are used as wavelength (de)multiplexing points and the design of ONUs (Figure 3a). In WDM-PONs as a basic multiplexing stage most frequently, a wavelength routing device (like an Arrayed Waveguide AWG, used in a single or in multiple stages) or wavelength filters with specific properties also including power couplers are used. To achieve higher flexibility next-generation WDM-PONs should have the capabilities of flexible wavelength routing to sub-trees and in turn provide some sort of wavelength agility in the ONU side. Different degrees of flexibility in wavelength routing could be provided by supporting sharing of wavelength across different sub-trees. The latter option assumes some sort of wavelength agility on the ONU side to operate in different wavelengths.

In [9][10], an efficient long-reach access network supporting the above features based on the introduction of the Active Remote Node (ARN) in the access network exploiting a Fibre-to-the-Curb (FTTC) architecture has been presented. The ARN is physically sited close to the end-user, e.g., where the Cabinet/DSLAM is located and its role is to terminate and (de)aggregate the traffic from/to a group of end-user ONUs in the same neighbourhood (Figure 3b). Thus, the ARN is the place where optical transparency is terminated creating a two-stage optical access network.



(a)

(b)

Figure 3. Generic WDM-PON architecture (a) and two-stage hybrid TDM/WDM PON

This architecture has been shown to reduce cost and increase scalability [10]. The most important features of the two-stage hybrid TDM/WDM PON of Figure 3b is that access rates beyond 10Gb/s per user can be achieved using cost efficient ONUs (e.g., XGPON [2]) while increasing aggregate capacities over the WDM trunk line.

Concluding, we summarize below the main points of the architecture described above. First, the flat hierarchy exploits optical transmission and multiplexing to a minimize the cost of packet processing and switching, since long-reach can be achieved (as proposed in [4][6]). The second contribution of this paper is that an integrated control plane architecture is proposed to implement this network hierarchy. Beyond simplification of the access routers and switches the proposed architecture can also lead to reduction of cost and power consumption due to the simplification of the Customer Premises Equipment (CPE), while maintaining very high access rates that can exploit mature 10Gb/s technologies. Below, we will finally show that the proposed integrated optical network architecture can exploit service transparency to also further reduce the implementation cost of multiple access control and access latency over the shared network paths of tree-shaped access networks.

IV. INTEGRATED RESOURCE RESERVATION SCHEME

The access architecture described above evidently defines two discrete resource reservation domains extending between the active nodes of the network (i.e., the customer ONUs, ARN and MEN), as depicted in Figure 2 and Figure 3b. Assuming TDM-PONs in the last drop from the ARN to the customer ONUs resource reservation in the first segment would be performed following the Medium Access Control (MAC) protocol and the associated traffic management mechanisms [1]-[3]. In both IEEE 802.3 based Ethernet PONs (EPONs) and the ITU based G-PONs this assumes the implementation of a number of discrete Classes of Service (CoS) and the corresponding per CoS queuing and request process in the ONUs and MAC scheduling of timeslot allocations per ONU and CoS (transmission grants) at the OLT [11]. If the two segments operate independently, forwarding of traffic aggregated by each OLT over the core network would then typically require routing/switching of traffic to output interfaces. Implementing appropriate output queuing policies CoS scheduling would handle the transmission over each output interface. This is schematically shown in Figure 4a, where M input (access) interfaces (OLTs) aggregating traffic from PONs with a split ratio N are interconnected to L core network destination nodes (Di) over K output (core) interfaces.

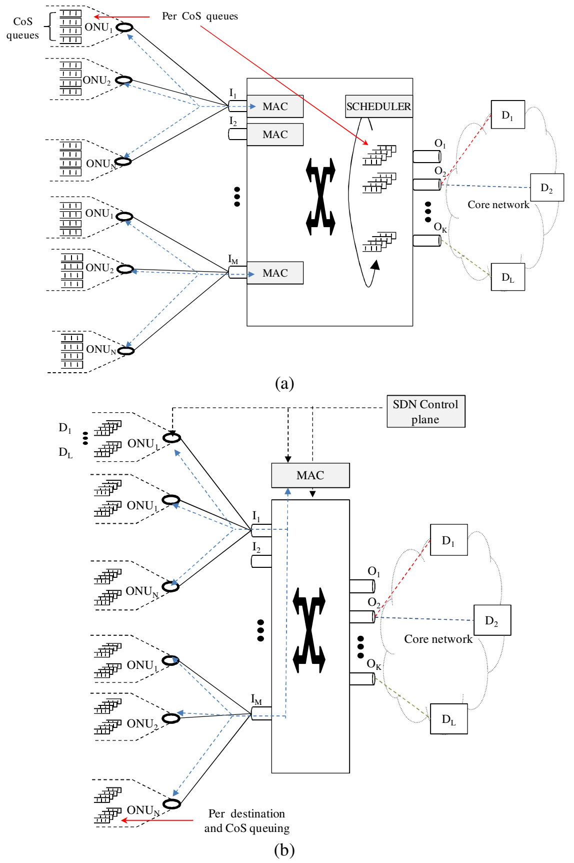


Figure 4. Distributed per segment queuing and scheduling (a) vs. SDN controlled distributed multiplexing and centralized MAC (b)

Implementing the SDN approach described in Section II, flexible flow association rules can be defined (e.g., using the OpenFlow protocol) across the entire optical access and core network up to the ONUs. This can make possible an alternative queuing and reservation policy (following the approach of [11]), which can handle collectively upstream requests and perform scheduling and traffic forwarding in a centralized manner, as shown in Figure 4b.

In this paper, we extend the work presented by Orphanoudakis et al. [11] to address hybrid WDM/TDM networks as proposed by Matrakidis et al. [9] assuming an SDN compatible mode of operation estabishing provisioned links over the core network to implement flow forwarding based on programmable rules. Thus, even packet switching at the ARN side can be simplified by appropriate frame tagging techniques like VLANs or MPLS. Finally, the centralized arbitration eliminates the requirement for buffering at the output interfaces.

It is worth noting that this scheme does not increase the complexity of ONUs, since per flow queuing is already implemented and only the number of queues is increased (not affecting the total memory requirements though, as will be shown next) to implement per destination queuing as determined by SDN rules. On the contrary, a single MAC entity can be used to collectively schedule upstream transmissions from all access networks so as to optimally utilize connections over the core network eliminating the need for output queues and scheduling. Thus, the network will operate as a distributed switch under the control of the centralized MAC arbiter. Since the single MAC entity has global knowledge of traffic conditions at each ONU as well as the status and service rate of each core interface we will show that improved access latency as well as throughput can be achieved.

V. Performance Evaluation

In order to evaluate the performance of the proposed scheme, a simulation model implementing the two-stage hybrid TDM/WDM PON shown in Figure 3b was developed using the OPNET simulation tool. The simulated topology comprises 4 PONs, with 16 ONUs each, the capacity of each PON being set to 10Gb/s. We consider two traffic classes (high priority and low priority class). The traffic arrival pattern of the high priority class were simulated by constant bit rate sources generating short fixed-size packets periodically. This service class is expected to serve mostly traffic from real-time services with high QoS requirements. The traffic arrival pattern of the low priority class were simulated by ON-OFF sources (modelling self-similar Internet traffic) with a burstiness factor of 8. Regarding the packet size, the tri-modal distribution was used as is proved to be good approximation of IP applications originating in Ethernet networks. In detail, it consists of packet sizes of 64, 500, 1500 bytes appearing with probability of 0.6, 0.2, and 0.2 respectively. The traffic mix included on average 30% high priority traffic and 70% low priority traffic. The number of destination core nodes (L) was set to 13. In order to simulate an unbalanced traffic demand per destination all packets from the first ONU in each PON, were set to have as destination a single core node, while for the traffic from all the other ONUs its destination was randomly selected by using a uniform distribution among all destinations.

Following the methodology of [11], we compared this architecture with a non-integrated one, where the OLT MAC schedules the transmission of each PON independently. In this architecture the OLT decides on scheduling by taking into account only the requirement for fair sharing of resources among all competing ONUs (denoted hereafter as “Distributed” MAC implementation referring to Figure 4a).

Figure 5 illustrates the probability density function (PDF) of access delay (across all queues of all ONUs) for a total offered load of 95%. Evidently, the proposed centralized architecture achieves lower aggregate access delay at high loads (at low loads demands per output interface remain low; hence, there is no impact of the MAC scheduling).

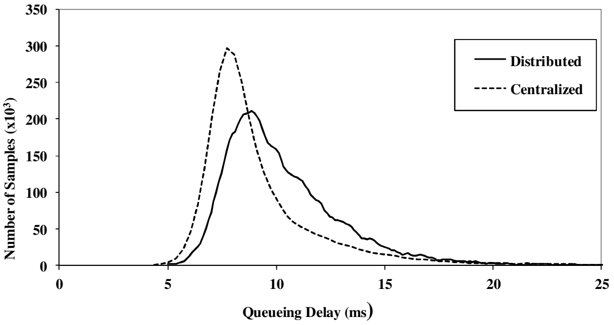


Figure 5. Probability Density Function (PDF) of access delay at 95% load.

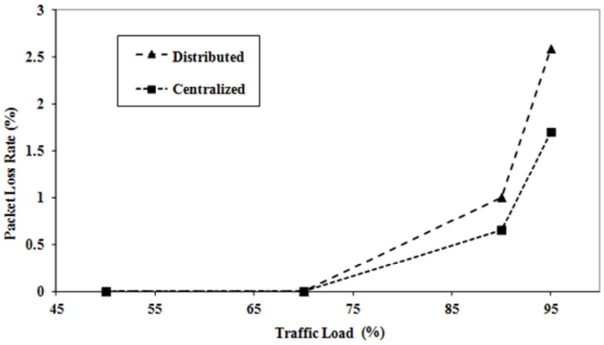


Figure 6. Packet Loss Rate (PLR due to buffer overflows.

The improved access latency of the centralized MAC we propose in this paper, also has an impact on the total buffering requirements as observed in Figure 6, where the packet loss probability in ONU queues is shown. As can be deduced by observing the results in Figure 6, the distributed MAC results in higher losses and reduced throughput for an ONU buffer size value of 16Mbits, while for the centralized MAC proposed here this value is up to 30% improved.

VI. Conclusions

We described an end-to-end optical network architecture that can efficiently scale to support the requirements of next generation networks. The proposed architecture exploits flexible core networks based on efficient traffic aggregation and optical switching and integrated access-core networks based on hierarchical long reach hybrid TDM/WDM PONs allowing access rates beyond 10Gb/s. Based on this network architecture, we proposed a unified control plane architecture that can perform end-to-end coordination and resource reservation based on SDN principles. Hence, ultra high capacity service transparent optical networks can be deployed exploiting the emerging SDN principles to implement traffic aggregation and flow switching. A scenario of operation of this integrated resource reservation scheme and centralized control was assessed by means of simulation and was shown to achieve improved performance in terms of throughput and access latency. Additionally, the proposed scheme can lead to lower complexity of the access nodes since a centralized MAC engine guided by an SDN control plane can replace multiple engines that would be needed in the case of distributed control.

Acknowledgment

This work has been co-financed by the EU (ESF) & Greek funds through the “Education and life-long learning” of the NSRF-Research Funding Program: ARCHIMEDES III “Investing in Knowledge Society through ESF”.

References

[1] ITU-T G.984.x “Gigabit-Capable Passive Optical Networks (GPON)”.

[2] ITU-T G.987.x, “10-Gigabit-Capable Passive Optical Networks (XG-PON)”.

[3] IEEE Stds. 802.3ah-2004, IEEE Stds. 802.3av.

[4] G. Weichenberg, V. W. S. Chan and M. Médard, “Design and Analysis of Optically Flow Switched Networks,” IEEE/OSA Journal on Optical Communications and Networking, vol. 1, no. 3, pp. B81-B97, Aug. 2009

[5] A. Stavdas, C. Matrakidis, C.T. Politi, T. Orphanoudakis, and J. Dunne, D. Chiaroni: “Optical Packet Add/Drop Multiplexers for packet ring networks” Th.2.E.1, vol. 4 – 103, ECOC 2008, 21-25 September 2008, Brussels, Belgium 2008.

[6] J. D. Angelopoulos, et. al., “An Optical Network Architecture with Distributed Switching Inside Node Clusters Features Improved Loss, Efficiency and Cost”, IEEE Journal on Lightwave Technologies, vol. 25, no. 5, May 2007, pp. 1138-1146.

[7] Open Networking Foundation: “Software-Defined Networking: The New Norm for Networks” ONF White Paper April 13, 2012”, April 2012.

[8] OpenFlow Switch Specification Version 1.3.1 (Wire Protocol 0x04), September 6, 2012, OpenFlow 1.3.1.

[9] C. Matrakidis, T. Orphanoudakis, and A. Stavdas, “Performance Evaluation of a Truly Cost-Efficient, Scalable and Green Optical Access Network” Photonics in Switching 2012, Ajacio, France, 11-14 Sep. 2012.

[10] T.G. Orphanoudakis, C. Matrakidis, A. Stavdas, and H.-C. Leligou, “Exploiting state of the art WDM-PON technologies for building efficient FTTC networks”, 15th Int. Conf. on Transparent Opt. Netw. (ICTON), Cartagena, Spain, June 23-27, 2013.

[11] T. Orphanoudakis, H.-C. Leligou, E. Kosmatos, and A. Stavdas, “Future Internet infrastructure based on the transparent integration of access and core optical transport networks”, IEEE/OSA Journal of Optical Communications and Networking, vol. 1, no. 2, pp. A205–A218, 2009.

[12] Erik Dahlman, Stefan Parkvall, Johan Sköld "4G – LTE/LTE-Advanced for Mobile Broadband", Academic Press, 2011.

[13] ITU-T G.993.1, "Very high speed digital subscriber line transceivers", June 2004.

[14] ITU-T G.Sup50, "Overview of digital subscriber line Recommendations", November 2011.

[15] Cisco White paper, "Cisco Visual Networking Index: Forecast and Methodology, 2013–2018", June 2014.

Hybrid Optical Network with Traffic Classification and Switching Selection Scheme Based on Fuzzy Logic

Ana Carolina de Oliveira da Silva
Post Graduate Program in Electrical Engineering (PPGEE)
University of Brasília
Brasília, Brazil
anacarol.os88@hotmail.com

William Giozza
Communication Networks Group
University of Brasília
Brasília, Brazil
giozza@unb.br

Abstract—**This paper presents a performance evaluation study of an OCS/OBS hybrid optical network using a traffic classification and optical switching selector scheme based on fuzzy logic. The fuzzy logic scheme verifies the statistical parameters of the input traffic and selects the appropriate optical switching paradigm. The hybrid optical network performance is evaluated in terms of block probability and efficiency in the use of the network resources. Our results show that the hybrid optical networking approach, when submitted to non-uniform traffic, uses network resources in a more efficient way compared to an optical network implementing one unique optical switching paradigm.**

Keywords - Hybrid Optical Network; Optical Circuit Switching;Optical Burst Switching; Fuzzy Logic.

I. INTRODUCTION

With the maturity of Wavelength Division Multiplexing (WDM) technology, and the increasing bandwidth demand in telecommunications networking, researchers and engineers have been looking for optical networks solutions that use the maximum capacity offered by fiber optics [1-5].

Moreover, it has been observed that the type of traffic carried by modern telecommunications infrastructures, supporting the integration of services provided by the Internet, differs greatly from old Poisson traffic models. Traffic properties such as long-range dependence and self-similarity are present from local networks to large backbones, preventing the modeling of present traffic behavior by a Poisson process [6-11]. Therefore it seems important to consider traffic self-similarity characteristics - measured by the Hurst parameter - for planning and management of the network [7][12][13].

Optical networks can be classified according to three optical switching paradigms, with different implications in terms of complexity and performance: Optical Circuit Switching (OCS), Optical Burst Switching (OBS) and Optical Packet Switching (OPS) [1]. OCS has a simpler management method and works by establishing light paths ("optical circuits") between source and destination nodes; however, as well known, this switching paradigm is inefficient when subjected to high granularity traffic. OPS is an attractive option since it can transport Internet Protocol (IP) packets directly in the optical network. However, the absence of an equivalent to the Random Access Memory (RAM) in the optical domain and other technological barriers restrain the use of this optical switching paradigm in the implementation of optical networks in the near future. OBS scheme gathers input packets at the source node and sends them to the destination node as an optical burst. Therefore, OBS constitutes a compromise between OPS and OCS paradigms, facilitating the network management and improving the use of the available bandwidth [1][3].

In order to better adapt the characteristics of different types of optical switching to the diversity of the traffic presently carried by major telecommunications infrastructures (network cores or backbones), the concept of hybrid optical network emerged, characterized by a single optical network implementing more than one of the optical switching paradigms. For instance, Gauger et al. [14] proposed a hybrid optical network OCS/OBS, aiming to take advantage of the high-granularity-traffic transport characteristic of OBS networks and the large bandwidth traffic supported by OCS networks.

This work presents a performance study of a hybrid optical network OCS/OBS, where the decision of the appropriate optical switching paradigm to incoming traffic is done on the basis of their statistical parameters using fuzzy logic.

This paper is organized as follows: Section II presents the hybrid optical network topology used, Section III describes the fuzzy classifier used to select the more appropriate optical switching paradigm to an incoming traffic pattern, Section IV presents the performance evaluation study and Section V presents the conclusions.

II. HYBRID NETWORK OBS/OCS

The OCS model adopted in this work uses First Fit as Routing and Wavelength Assignment (RWA) algorithm [2]. The wavelength assignment strategy assumes that network nodes do not perform wavelength conversion.

The OBS model was implemented according to the Just Enough Time (JET) protocol [1], assuming the existence of a parallel network of infinite capacity to the transport of control packets associated to the optical bursts. Therefore, resources are busy end-to-end, simulating the resource reservation, and unoccupied as the optical bursts pass through each link. The validation of the OBS implementation was made by comparison with results presented in Teng and Rouskas [16].

The OCS/OBS hybrid optical network includes a traffic classifier and switching selector module based on fuzzy logic in each edge node. Figure 1 illustrates the OCS/OBS hybrid optical network model used in this work.

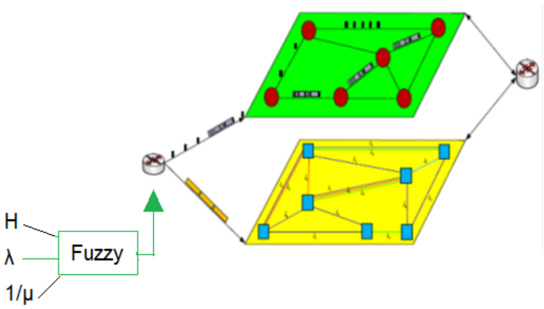


Figure 1. Parallel hybrid optical network with traffic classifier.

The traffic classifier and optical switching selector module works based on statistical parameters of the incoming traffic - packet arrival rate (λ), inter-arrival time (Δt) and Hurst parameter (H) and selects the more appropriate optical switching paradigm (OCS or OBS) to apply to the incoming traffic at each time slot, at each edge node of the parallel hybrid optical network (Figure 1).

III. TRAFFIC CLASSIFIER

Fuzzy logic was chosen to implement the traffic classifier and switching selector (in short, the traffic classifier) because of its robustness in handling imprecise input values and its flexibility to work with traffic characteristics evolution. In general, problems solved by fuzzy logic can be solved by other tools; however, in the case of nonlinear problems, fuzzy logic presents advantages, avoiding the creation of complex mathematical models and requiring only the knowledge of an expert. Once the traffic classifier handles different statistics that cannot be added, it can be considered as a non-linear system.

Fuzzy analysis divides the problem into three steps [17]:

- Fuzzyfication – The input variables are transformed into linguistic values associated with an inference function;
- Inference – The linguistic values are analyzed by an array of rules;
- Defuzzification – Linguistic output of the rules matrix is transformed into a numeric value.

In this work, the membership functions of the inputs and outputs of the traffic classifier were constructed considering the expected network behavior and by analyzing traffic traces available in the literature [6][10][11][18][21]. The fuzzyfication step was performed from the numerical values corresponding to the statistical characteristics of the traffic. The traffic classifier was developed based on the MATLAB fuzzy toolbox.

The input λ corresponding to the arrival rate of packets in the network, illustrated in Figure 2, considers Low arrival rates lower than $4x10^5$ packets/s; Medium rates between $3x10^5$ packets/s and $5x10^5$ packets/s; and High rates superior to $4x10^5$ packets/s.

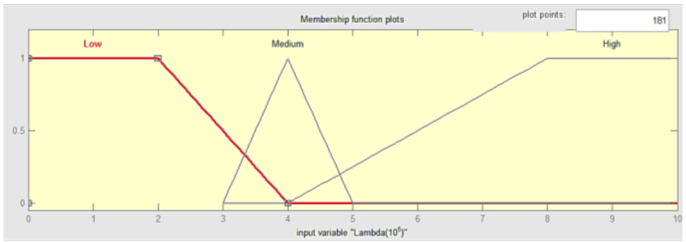


Figure 2. Traffic classifier – input λ.

The input Δt corresponding to the inter-arrival time (Figure 3) considers Low, intervals shorter than 3μs; Medium intervals between 2μs and 4μs; and High intervals longer than 3.5 μs.

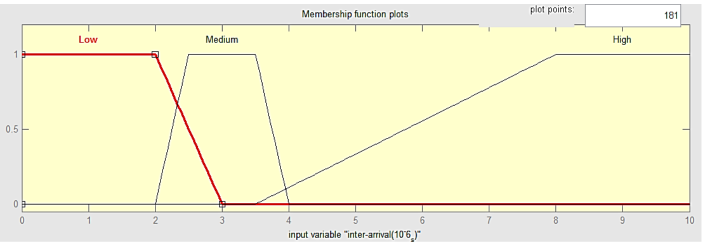


Figure 3. Traffic classifier – input Δt.

The input H corresponding to the Hurst parameter (Figure 4) considers Poisson traffic values lower than 0.55; Hybrid, a trapezoidal function from 0.4 to 0.6; and Self-similar for higher values than 0.55.

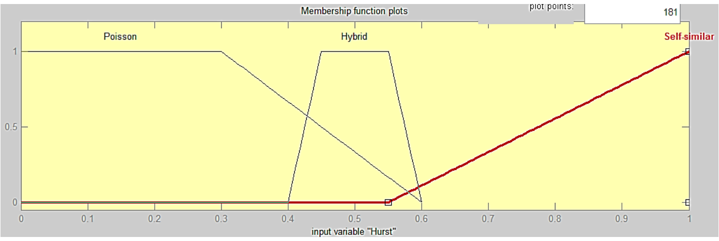


Figure 4. Traffic classifier – input H.

The output of the traffic classifier is adimensional and is on a scale 0-1 as shown in Figure 5. For output values between 0 and 0.7, the incoming traffic is considered appropriate to OBS transport and for values greater than 0.5, the incoming traffic is considered appropriate to OCS transport. These values were selected in a way that the centroid of the resulting area after the defuzzification tends to the right side, implying higher priority to Quality of Service (QoS) than to the economy of resources.

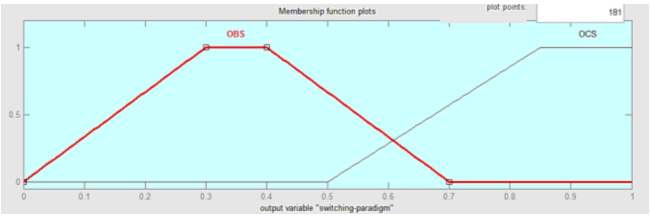


Figure 5. Traffic classifier –output.

The matrix rules (fuzzy inference step) for the traffic classifier is described in Tables 1 to 3.

TABLE I. MATRIX RULES FOR THE CLASSIFIER – ΔT LOW.

Hurst Parameter (H)	**Packet Arrival Rate (λ)**		
	Low	***Medium***	***High***
Poisson	OBS	OCS	OCS
Hybrid	OCS	OCS	OCS
Self-similar	OBS	OCS	OCS

TABLE II. MATRIX RULES FOR THE CLASSIFIER – ΔT MEDIUM.

Hurst Parameter (H)	**Packet Arrival Rate (λ)**		
	Low	***Medium***	***High***
Poisson	OBS	OCS	OCS
Hybrid	OBS	OCS	OCS
Self-similar	OBS	OBS	OCS

TABLE III. MATRIX RULES FOR THE CLASSIFIER – ΔT HIGH.

Hurst Parameter (H)	**Packet Arrival Rate (λ)**		
	Low	***Medium***	***High***
Poisson	OPS	OBS	OCS
Hybrid	OPS	OBS	OBS
Self-similar	OBS	OBS	OCS

The defuzzification process is done using the centroid method [17], which provides as output the center of gravity of the fuzzy set. Figure 6 illustrates a temporal series of Voice over IP (VoIP) traffic used to validate the traffic classifier.

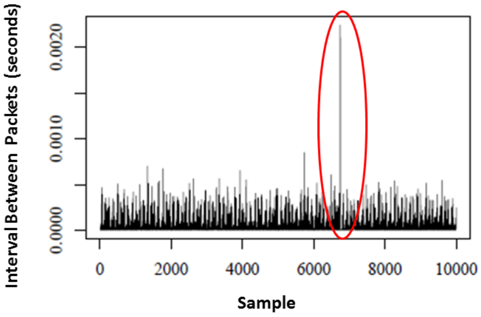


Figure 6. Temporal Serie of a VoIP traffic [11].

The highlighted moment in Figure 6 correspond to an instant which could cause a wrong classification, since it is a high inter-arrival time. However, the use of fuzzy logic to the classification allows considering the statistical parameters together, avoiding a wrong decision, especially in the presence of spurious in the incoming traffic, making the traffic classifier more robust. The traffic classifier output obtained for the input illustrated in Figure 6 was OCS which is consistent, since VoIP is a service which requires high QoS.

IV. PERFORMANCE EVALUATION

The simulation tool used in this work is an extension of the Transparent Optical Network Simulator (TONetS) presented in Soares et al. [15] allowing performance comparison between a parallel hybrid optical network as in Gauger et al. [14] and an optical network OCS under the same traffic conditions. TONetS is an educational simulation tool, developed in JAVA and conceived in blocks to allow its evolution. TONetS, initially developed to work with OCS switching, has been adapted in this work to support OBS switching in a parallel hybrid optical network.

The main gain when using hybrid optical networks is the economy and the optimization of resources while transporting sparse traffic. In order to assess these gains, we define the metric Relative Resource Economy:

$$E_U(\%) = \frac{U_{OCS}-U_{OBS}}{U_{OCS}} \times 100 \qquad (1)$$

where U_{OCS} and U_{OBS} are the utilization of the network using OCS and OBS, respectively.

In order to evaluate the performance difference between the OCS and OBS switching paradigms and the hybrid optical network approach proposed, a simple topology with 3 nodes was tested, using the three forms of switching (OBS, OCS and Hybrid). Each simulation considered: 1000 requests, two replications and a confidence level of 0.95. Forty wavelengths by optical fiber link were assigned using the First Fit algorithm and all simulations were performed considering the optical network nodes without wavelength conversion capability.

Using the simple 3-node topology illustrated in Figure 7, the OBS and OCS schemes were evaluated separately and together submitted to an incoming non-uniform traffic handled by the traffic classifier in order to select which one is more appropriate. In the case of the hybrid optical network approach, the Hurst parameter was varied uniformly between 0.3 and 1.0 to ensure the dynamism in the network. In all scenarios analyzed, 40 optical channels were used (i.e., wavelengths) by link subjected to an initial load of 10 Erlangs and 140 Erlangs, with increments of 10 Erlangs, in order to compare performances based on metrics such as Blocking Probability and Utilization.

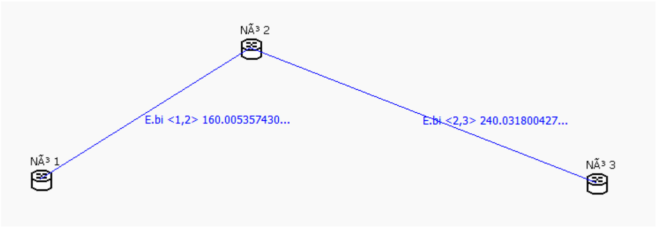


Figure 7. Analysed topology.

As a first scenario studied, an OCS network with the topology illustrated in Figure 7 was submitted to an initial load of 10 Erlangs with five increments of 10 Erlangs. Figure 8 illustrates the overall OCS utilization in this case.

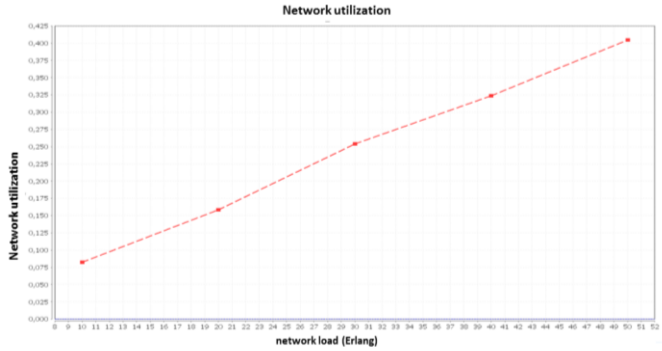


Figure 8. Network utilization with OCS and initial load of 10 Erlangs.

Due to the low load level, a blocking probability almost negligible was observed and a low utilization of the network resources of the OCS network. Similar results were observed with an OBS network and a hybrid OCS/OBS network. However, OBS switching achieved a slightly lower performance, for instance, 0.15% of blocking probability at 50 Erlangs.

As a second scenario studied, the same OCS network previously presented (Figure 7) was subjected to an initial load of 140 Erlangs with five increments of 10 Erlangs. Figures 9 to 11 illustrate the general network utilization, the blocking probability, and wavelength utilization, respectively.

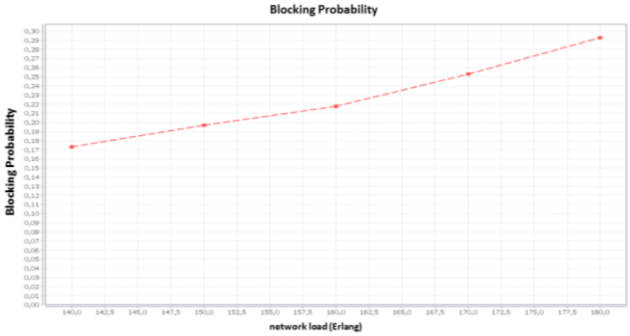


Figure 9. Blocking Probability with OCS and initial load of 140 Erlangs.

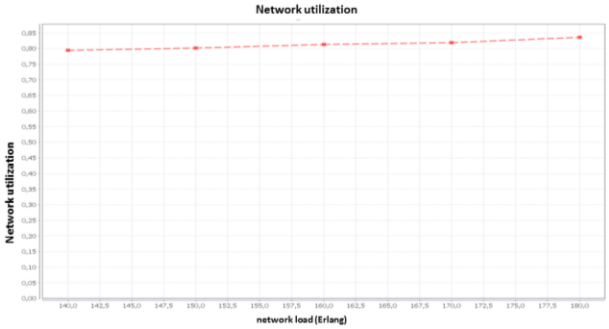


Figure 10. General network utilization with OCS and initial load of 140 Erlangs.

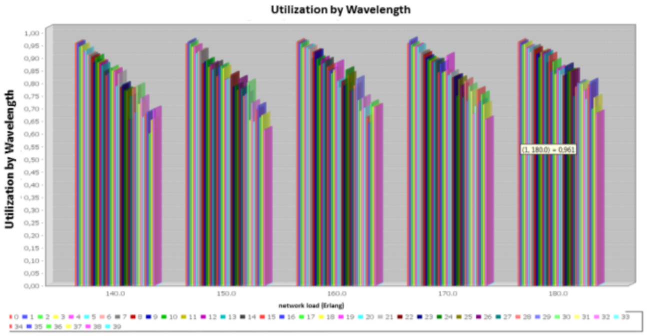


Figure 11. Utilization by Wavelength with OCS and initial load of 140 Erlangs.

In this second scenario, the blocking probability observed was high - 29.2% at 180 Erlangs - as well as the network utilization - 83.6% capacity at 180 Erlangs. The same traffic load condition was applied to OBS switching, resulting in the utilization by wavelength performance shown in Figure 12.

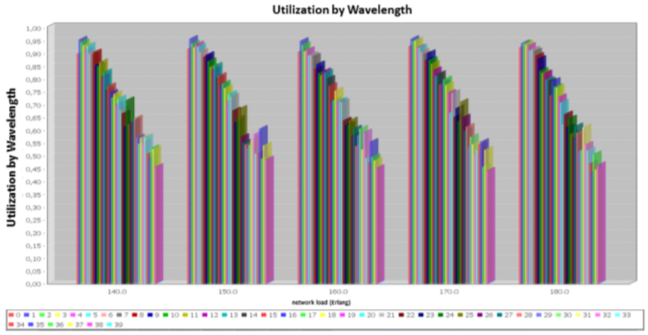


Figure 12. Utilization by Wavelength with OBS and initial load of 140 Erlangs.

From Figures 11 and 12, it can be observed that OBS switching allows improving the network performance. This result is consistent and can be associated with the fact that with OBS, intermediate links are released after the passage of the optical bursts. Therefore, these resources are available to new requests. Moreover, when the network load is high, intermediate links becomes relevant to performance. Table 4 illustrates the Relative Resource Economy, as in (1), obtained using the OBS paradigm.

TABLE IV. RELATIVE RESOURCE ECONOMY

LOAD (E)	OCS	OBS	$E_U(\%)$
140	0.802	0.694	13.47
150	0.801	0.715	10.74
160	0.822	0.714	13.14
170	0.822	0.709	13.75
180	0.83	0.726	12.53

As expected, OBS showed better performance in terms of resource economy. For all load conditions, we obtained savings higher than 10% compared to OCS; this is an important factor especially when considering large networks which increase the availability of resources requiring large investments.

The same load condition of the previous scenario was applied to the hybrid OCS/OBS network, with the value of the Hurst parameter ranging uniformly between 0.3 and 1.0. The performance obtained in terms of utilization by wavelength is shown in Figure 13.

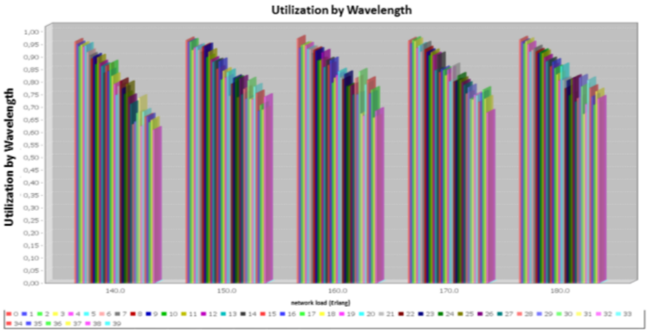


Figure 13. Utilization by Wavelength with Hybrid OCS/OBS switching and initial load of 140 Erlangs.

The hybrid OCS/OBS switching, as expected, showed intermediate results between those obtained with the OCS and OBS separately. While hybrid OCS/OBS network presented smaller network utilization in the presence of low granularity traffic, OCS network appeared to be transparent to the granularity of traffic, wasting the available resources. The OCS/OBS blocking probability was about 5% lower than OCS at 180 Erlangs. These performance differences are more apparent according to the incoming traffic. In the case of sparse traffic, a larger portion of the traffic is switched by bursts, making the overall performance of the OCS/OBS network closer to that obtained with the OBS paradigm. For more continuous traffic, the opposite occurs. Recent studies [18-21] indicates that backbones traffic is mainly composed of data packets, which implies that a significant portion of the traffic can be switched by bursts, optimizing the existing infrastructure and avoiding the waste of resources.

V. CONCLUSIONS AND FUTURE WORK

This work presents a performance study comparing an OBS/OCS hybrid optical network approach to the OCS and OBS paradigms isolated.

The OCS/OBS hybrid optical network studied uses a traffic classification and optical switching selector scheme based on fuzzy logic.

Our results show that, when the traffic is non-uniform, a hybrid optical network approach uses the available resources in a more efficient way than an optical network implementing one unique optical switching paradigm.

The use of Fuzzy Logic at each edge hybrid node makes traffic classification scheme robust to errors in the estimation of the input parameters, traffic growth and the emergence of new applications. The use of statistical traffic parameters to select the optical switching paradigm makes this scheme transparent to protocols. In the work presented by Lee [23], the author proposes a hybrid network and suggests a classification of traffic according to the duration of the flow. The decision threshold, however, remains an open issue. The classifier presented here fills this gap, consisting of a generic and robust solution, which can be seen as an offshoot of the work of Lee.

This work brings some challenges to be investigated, such as the evaluation of reducing latency in high QoS services using hybrid optical networks and the consequent reduction in dispute over resources with lower QoS services and the use of fuzzy logic in other ways to manage traffic.

REFERENCES

[1] J. P. Jue, V. M. Vokkarane,“Optical Burst Switched Networks”. Springer, 2005.

[2] R. Ramaswami, K. Sivarajan, and G. Sasaki, “Optical Networks – a practical perspective”, 3rd. ed., Morgan Kaufmann, 2010.

[3] D. Tafani, C. McArdle, and L. P. Barry, “Analytical Model of optical burst switched networks with share-per-node buffers”, IEEE Symposium on Computers and Communications – ISCC, Jun. 2011, pp. 512-518.

[4] M. J. O'mahony, et al. “Future Optical Networks”. Journal Lightwave Technologies 24, Dec , 2006, pp. 4684-4696.

[5] G. Corazza, W. Cerroni, G. Leli, C. Raffaelli, M. Savi, and N. Stol, “Analitical Model of 3-level QoS Schedulig in Hybrid Optical Networks”, International Conference on Computing, Networking and Communications (ICNC) Workshop on Computing, Networking and Communications – IEEE, Jan. 2013, pp.180-184.

[6] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson “On the self-similar nature os Ethernet traffic”. SIGCOMM ’93, Vol. 23, Oct. 1993, pp. 183-193.

[7] V. Paxson, S. Floyd, “Wide-Area Traffic: The Failure of Poisson Modeling”. IEEE/ACM Transaction on Networking, Jun. 1995, pp. 226-244.

[8] M. E. Crovella, A. Bestavros, “Self-similarity in Word Wide Web: evidence and possible causes”, IEEE Transactions on Networking, Vol. 5, no. 6, Dec. 1997, pp. 835-846.

[9] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, “Long range dependence in Variable Bit Rate Video traffic”, IEEE Transactions on Communications, Vol. 43, Apr. 1995, pp. 1566-1579.

[10] C. Park, F. Hernandez-Campos, J. S. Marron and F. D. Smith, “Long Range Dependence in changing internet traffic mix”. Computer Networks 48, Jun. 2005, pp. 401-422.

[11] C. M. Pedroso, J. P. Caldeira, and K. Fonseca, “Caracterization of VoIP traffic”, XVI Seminário de Iniciação Científica e X Mostra de Pesquisa Pontífica Universidade Católica do Paraná, Nov. 2008 (available in Portuguese).

[12] L. P. R. Kumar, S. K. Kumar, D. M. Reddy, and M. R. Perati, “Analytical model for performance study of the switch under self-similar variable length packet traffic”, Proceedings of the World Congress on Engineering and Computer Science, Vol. 1, Oct. 2010.

[13] T. Karagiannis, M. Molle, and M. Faloutsos, “Long- range dependence – Ten years of Internet traffic modelling”, IEEE Internet Computing, Vol.5/8, Sep. 2004, pp. 57-64.

[14] C. M. Gauger, et al. – “Hybrid Optical Network Architectures: Bringing Packets and Circuits Together”. IEEE Communications Magazine, Vol. 44, n. 8, 2006, pp. 36-42.

[15] A. Soares, G. Durães, W. Giozza, and P. Cunha, “TONetS: a performance tool of transparent optical networks”, Proc. of XXVII Sistemas Computacionais e de Computação, Jun. 2007, pp. 579-594. (available in Portuguese).

[16] J. Teng, G. N. Rouskas, “A comparison of the JIT, JET, and horizon wavelength reservation schemes on a single OBS node”, Proceedings of the First International Workshop on Optical Burst Switching, 2003.

[17] E. Cox, The fuzzy systems handbook: a practitioner’s guide to building, using, and maintaining fuzzy systems . New York: AP Professional, 1994.

[18] http://www.caida.org/data/passive/passive_trace_statistics.xml, [retrieved March, 2014]

[19] Y. Won, R. Fontugne, K. Cho, H. Esaki, and K. Fukuda, “Nine years of observing traffic anomalies: trending analysis in backbone networks”, International Symposium on Integrated Network Management (IM 2013), May 2013, pp. 636-642.

[20] http://www.fukuda-lab.org/mawilab/, [retrieved March, 2014]

[21] http://www.cs.columbia.edu/~hgs/internet/traces.html, [retrieved March 2014]

[22] C. Xin, C. Qiao, Y. Ye, and S. E. Dixit, – “A hybrid optical switching approach”. IEEE GLOBECOM 2003, Dec. 2003, pp. 3808-3812.

[23] G. M. Lee, “Optical hybrid switching with flow classification in IP over optical network”, PhD thesis, Korea Advanced Institute of Science and Technology, School of Engineering, 2007.

Performance of Low Buffer Resource Flexible Router for NoCs

Israel Mendonça dos Santos
and Felipe M. G. França
PESC/COPPE, Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brasil
Email: {israel, felipe}@cos.ufrj.br

Victor Goulart
E-JUST Center, Kyushu University
Fukuoka, Japan
Email: goulart@ejust.kyushu-u.ac.jp

***Abstract*—Nowadays, the performance of advanced multi-core systems is mostly limited by communication bottlenecks instead of computing speed or memory resources. Interconnection networks start to replace buses as the standard system-level interconnection infrastructure. Many researchers have been working on on-chip interconnection networks (Network-on-Chip - NoC) to improve its performance in terms of latency, throughput and power consumption. Buffers are the most influential element affecting performance in this architecture, and also the most expensive resource. The vast range of applications concurrently executing in the network and their different communication patterns leave some buffers of the routers underutilized. In this paper, we study the performance of low buffer resource flexible routers which can adaptively schedule resources (buffers) according to the dynamic behavior of the communication demand in the NoC. Our simulations showed this router architecture was able to improve throughput up to 21% or have similar performance while using half the number of buffers compared to a standard router.**

Keywords—Flexible Routers, Network-on-Chip (NoC), Adaptive Router, Buffer Resources.

I. Introduction

Processor technology scaling down allowed the integration of billions of transistors on a single chip, and the emergence of multi-core systems with dozens of hundreds of processors and multiple (distributed) memories. Such high complexity systems on chip are hard to implement with traditional bus-based communication infrastructure. To solve this problem the concept, of Network-on-Chip (NoC) [1] [2] [3] was introduced and is the focus of recent research. In a NoC-based system, on-chip modules exchange code or data using a network as a sub-system for the information traffic. NoCs are built from multiple point-to-point data links interconnected by routers. Recently, researchers have been pushed to substitute the bus architectures by NoCs due to their scalability and efficiency benefits. In the design of NoCs, high throughput and low latency are both important design parameters and the router is the essential key to meet these requirements. High-throughput routers are required to allow an outflow of packets that matches the needs of the applications. Normally, a higher number of buffers improves performance (network throughput), but they also impact the power consumption, being responsible for about 64% of the total router leakage power [4].

NoCs can have different communication performance and costs, depending on their architecture features and design, and their target applications. Designing the same NoC router to cover the whole spectra of applications would lead to an oversize and expensive router, in terms of area and power. On the other hand, if one tries to design an application specific router for different markets, the designer would need to take several design decisions in a very early stage, eventually compromising scalability and optimization capabilities for distinct application behaviors.

According to Dally and Towles [5], a router role lies basically in efficiently traversing packets through the network links. Router buffering is used to hold packets that are unable to be forwarded to the desired output port due to contention. An ideal router should be able to adapt to different application requirements at runtime without compromising its performance. According to the experiments realized by Xuning and Peh [4], even at high loads, there were still 85% of idle buffers, which highlights the importance of performance optimization. In fact, it has been observed that storing a packet consumes far more energy than transmitting if forward [6].

Given the aforementioned significance of the NoC buffers, the concept of Flexible Router [8] [9] was introduced; it has a new router architecture approach that fully utilizes the available buffers in a flexible or adaptive way. When all the buffers of an input port are already in use and the upstream router asks for a buffer space, instead of just denying the request, and so causing contention, the flexible router will try to allocate any other empty buffer to the upstream, and, by doing so, avoiding contention.

Flexible Routers are a potentially promising architecture approach to overcome buffer under utilization and improve performance. In this work, a router that can reconfigure the buffers to better attend the communication demands is analyzed under various architecture design parameters. The result showed that the proposed router improved over the traditional router's performance in terms of throughput up to 21% and having similar performance when using half of the buffers.

This paper is organized as follows. Section II gives background information about the traditional baseline router and summarizes the related works. The Flexible Router architecture and functionality is explained in Section III. Section IV shows the experimental framework and results. Finally, the conclusion is shown in Section V.

ISBN: 978-1-61208-368-1

II. Background and Related Work

Before explaining the functionality of the flexible router and its advantages, we present a brief introduction about the operation of the base router and the state-of-art concerning it.

A. Baseline Router

The baseline router used in this paper uses the architecture proposed by Dally and Towles [5], Figure 1. It is composed of several components that collectively implement the routing function and flow control functions. These functions are required to store and forward flits en route to their destinations through the network. The components are input controllers, one or more buffers per input or output port, a routing component, an arbitration component, and output controllers. A base router usually implements wormhole flow-control [10] with P ports, where P depends on the dimension of the topology. In a 2-D topology, P=5, as it includes the 4 ports to the neighbor (North (N), South (S), East (E), and West (W)) and the local port (L) (from/to the processor or memory core). Every input port has a certain amount of buffers to store flits. Flits are saved in buffers because they were unable to cross the network due to diverse factors, such as lack of credits on the link, lose the switch arbitration, among other things. The buffers are organized as *First In First Out* (FIFO) queues. Virtual Channel flow control [11] is the technique that separates the allocation of the buffers from the allocation of the channel. In case of blocking in one buffer, the input port can make use of another buffer. Flits entering the router are put in one of these buffers, called Virtual Channels (VCs), depending on their Virtual Channel Identifier (VCID). Queues on the input port are connected to a crossbar that links any input buffer to any output port.

The packet processing on the base router can be separated in a five-stage pipeline: The first stage is the Routing Computation (RC), where the Routing Module will decide an output port based on information extracted from the packet header. After RC, Virtual Channel Arbitration (VA) is done. During this stage, a buffer on the downstream router is selected. The VC Allocator sends a request signal to the downstream router that replies it with a grant signal containing a VCID (or a ”no-buffer signal” depending of the availability of the buffer). Based on the answer of the downstream router, the requester upstream router will attach the replied VCID to every flit of the packet so that when it arrives in the downstream, it is put in the correct virtual channel. After the VA stage, we have the Switch Arbitration stage (SA). In this stage, the flits will compete for the usage of the crossbar. At the Switch Traversal (ST) stage the flits cross the crossbar in the direction to the output port. And, finally at the Link Traversal (LT) stage, the packet is forwarded to the next router.

The main limitation during the Base Router operation is the contention problem. Contention occurs when a buffer request at some input port is blocked because this port has no buffer to allocate this packet. Such contention may lead to further blockings in the network and hence congestion occurs [12] [13], which degrades the performance of the network.

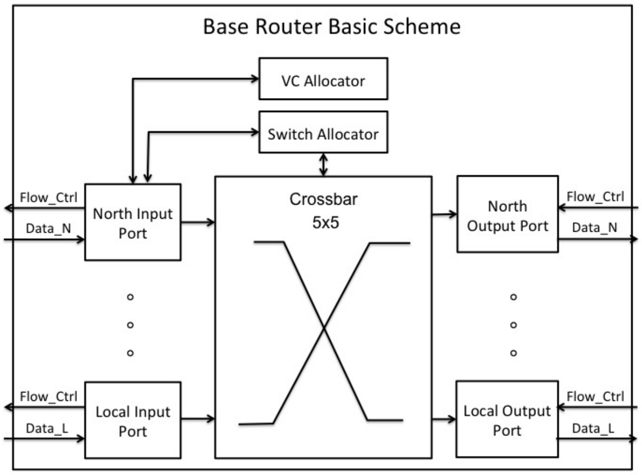


Fig. 1. Base Router Architecture [5].

B. Related Work

Several works have been proposed to maximize the performance through different buffer management techniques within the router. In Karol et al. [14] and Ramanujam et al. [15], a shared-buffer scheme is proposed to emulate an Output Buffered Router (OBR), but their mechanism greatly increases the number of the buffers of the router by adding extra Middle-Memories and a difficult control logic, which increases the area by 58% and the power usage by 35%.

In ViChaR [16], a Unified Buffer Structure (UBS) [17] is proposed with a dynamic virtual channel regulator called ViChaR, which also dynamically allocates virtual channels and buffers according to the network traffic conditions. They used the unified buffer unit instead of partitioned buffers as in our approach. The UBS is committed to create an illusion that the number, and size, of virtual channels of the router varies according to the traffic load. As the traffic load increases, ViChaR disposes more virtual channels to amortize the effect of the traffic. Although this mechanism has notable advantages, it has also suffered from a complex logic implementation increasing the area and probably causing power overheads.

In [18], Matos et al. propose the usage of a partitioned buffer that could lend/borrow buffer slots from each neighbor is proposed. By doing so, the depth of each channel could vary in time according to the traffic demands. But virtual channels were not used, only one buffer was available per channel, and since it could lend any buffer slot from its neighbors, it needed several multiplexers to provide the correct functionality of the mechanism, increasing the area and power demand. Also, the size of the multiplexer was determined by the size of the buffer, increasing linearly the necessary area. In our approach, the size of multiplexers is fixed, so the depth of the virtual channels have no influence on the size of our multiplexer. Also, our control mechanism is simpler, allowing the router to work in high frequencies, differently from the two previously mentioned approaches.

The idea of the flexible router was proposed in [8] and [9]. In this works, flexible router was implemented using Store-and-Forward (SAF) flow control and Virtual Channels were not used. Due to this limited implementations, it was not possible to verify all the improvements that the flexible router architecture could provide. In this work we do a broader evaluation under different design parameters for the router, such as Wormhole flow-control, and use of Virtual Channels, analyzed under different traffic patterns, packet sizes and number and size of virtual channels.

III. Proposed Flexible Router

Statistically, not all buffers are used all the time, leading to underutilization of these valuable resources. In our architecture, it is possible to dynamically allocate idle buffers to other channels. The proposed Flexible router addresses the poor utilization of buffers which lead to lower than the theoretical ideal throughput. At the same time, it tackles the area constraints of implementing a shared buffer router. With a simple control mechanism and the addition of a module called FIFO Flexibility Controller (FFC), the virtual channels are decoupled from the input port, as now any flit is able to acquire any buffer provided that VC is not in use already. Essentially, FFC virtually provides more buffers allocation resources between routers.

To guarantee packet delivery, a flow control mechanism needs to be employed. The proposed flexible router uses on/off flow control in order to prevent the drop of packets. If the downstream router is about to get full, a control signal is sent to the upstream router; by receiving this message, the upstream knows that the VCs of the downstream are about to get full and stop transmitting flits until another message arrives informing that the downstream router buffers are free.

Since power is a valuable resource in NoC domain, the power-performance trade-off of variable size and amount of VCs configurations must be explored. Although, for some cases, the performance of the flexible router can be the same as the baseline router, we try to reduce the amount of resources of our router and still have reasonable performance. Therefore, we evaluate configurations in which the flexible router has less buffer space available than the baseline router and still achieve similar results.

Finally, on-chip routers ought to operate at high clock frequencies. It is a design need to carefully architect the pipelines with low complexity logic at each stage. Our design employs a balanced five-stage pipeline. The proposed architecture can achieve higher performance than the traditional router with comparable buffering while adding reasonable area and complexity overhead in managing the flexibility of the input ports.

A. Flexible Micro-architecture and Pipeline

Figure 2 shows the router micro-architecture of the proposed flexible router and its corresponding five-stage pipeline. Incoming flits are first stored in their respective input buffer, which are segmented into a certain amount of VCs. The RC stage of the flexible router, differently from the base router, employs a look-ahead routing scheme, where the output of a packet is computed based on the destination coordinates, one hop in advance. This happens only for the head-flit of the packet. This is done so that the FFC mechanism can use this information to avoid deadlocks. Further explanation about deadlocks will be provided in subsection B. The VA stage is substantially different from the base router. Instead of arbitrating for free VCs in a specific input port at the downstream, it can be extended to the other buffers belonging to other input ports if they are not being used. After the VA stage, the router pipeline is exactly the same as the one on the baseline router. Subsequently, the packets dispute for the use of the crossbar on the SA stage. Winning flits of the SA stage traverse the crossbar at the ST stage, and finally the flit departs from the output port at the LT stage.

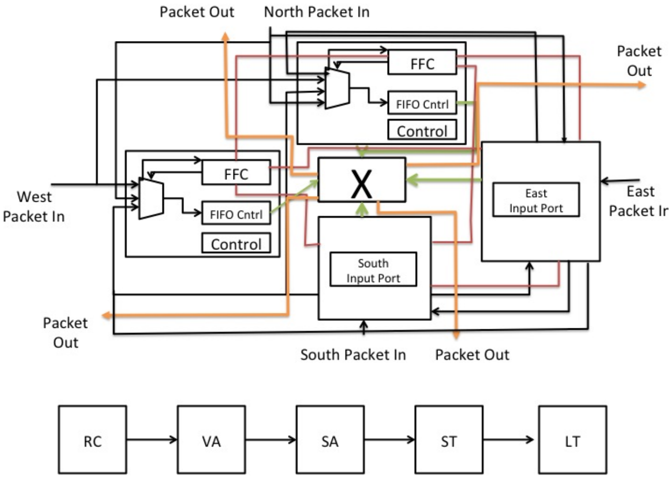


Fig. 2. Flexible Router Architecture.

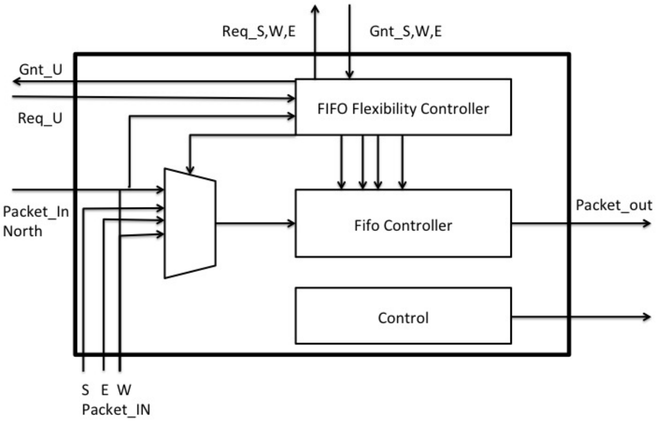


Fig. 3. Modified input port. A multiplexer and an extra module called FFC was employed.

Figure 3 shows proposed input port modules of the North Channel as an example. In this architecture, it is possible for a channel to lend its idle buffers to other channels. In conformity with this figure, each input port now has an extra multiplexer and an extra unity called FFC. The multiplexer is required in order to deviate the packets from their original channel and store them in the North buffer. The FIFO Flexibility Controller deals with the virtual channel allocation. Besides allocating VCs in its respective input port, it also works by allocating free

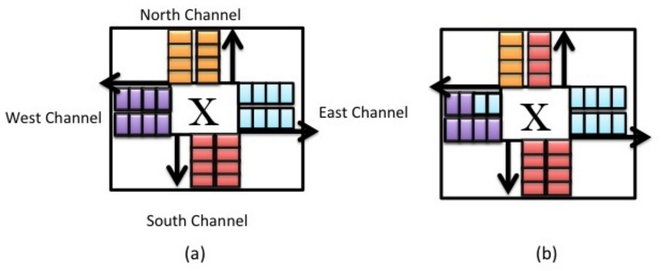


Fig. 4. (a) Baseline router designed with two buffers per port and FIFO depth 4; (b)Flexible router with the buffers reconfigured to attend the demand.

virtual channels that belong to other input ports. Whenever a header flit enters the VA stage at the upstream router and issues a request to the FFC of the downstream one, this request will be forwarded to the FFC in the downstream router. FFC will preferentially first look for free VCs into its corresponding input port, if it fails to do so, it will try to reserve a VC belonging to another input port. The process of reserving a VC at the neighbor input ports is an asynchronous process with three stages started whenever a FFC detects that it will not have enough buffer space to house the incoming flit. The first stage of the process starts when it issues a request to other FFCs. In the second stage, the FFC that receive the request message, will lookup into its virtual channels and if it finds an empty VC, a reply signal is sent with an ACK (if there are no free VCs, it will send a NACK). The last stage of the communication is when the issuer of the request message receives the replies from its neighbors. Then it makes a selection between the ACKs and replies the VCID to the upstream router (in case of receiving just NACKs, it replies a NOT AVAILABLE message).

To keep track of the availability of the VCs, we created an auxiliary structure called the Flexible Availability Table. The table contains the identification of each buffer and its status. If the FFC reserve one buffer, it is marked in the table as *reserved*. At the end of the usage of this buffer, the FFC changes the status in the table to *available*. This kind of structure reduces the amount of communication because the FFC asks directly to a port that is known to have empty VCs.

In this design, the usage of local input port buffers is not considered. The flexible router local port is not different from the one in the baseline router. Only the North, South, West, and East ports were changed. This is done in order to make the Flexible Router mechanism simple and not affect the injector queues.

The policy of choice of the neighbor VC, in case of more than one ACK, may influence the router performance. In our router, it was employed Dimension Order XY (DOR X-Y) algorithm [7], so the probability of filling up the buffers of the X direction is greater than the ones in the Y direction. Considering this fact, the Buffer Choosing Algorithm tries to choose buffers from ports that have less probability to be used through the execution of the system. The algorithm will be better clarified in subsection B.

Figure 4 shows how the VCs of the flexible router can be reconfigured. First, the depth of the buffers must be defined at design time, in this example, it has size equal to 4 and the packet size is 2, as illustrated in Figure 4(a). After this, the traffic is unbalanced and an entire virtual channel from North is lent to South port, and two buffer slots from West port are lent to East port, as shown in Figure 4(b). One can look that even though the West port houses a packet from the East port, it also can store packets from other ports as well. Flexible router has no issue about mixing packets from different ports in the same buffer, the only thing to be taken into consideration is not to interleave the packets. The mix is possible because once the packet acquires a VC, the routing information of the reader will be used for all the following packets of the VC until a tail flit is founded. If we interleave packets, body flits would go to wrong routes. The allocation of several packets in the same input port is possible because the tail flit releases the VC when it leaves the router. By doing so, even if the buffer is not empty, it can be reallocated to another packets if the previous packet is totally inside the buffer. We do not allow flits to go to different buffers because it would add complexity to organize these flits. The basic mechanism of the FIFO is thus kept.

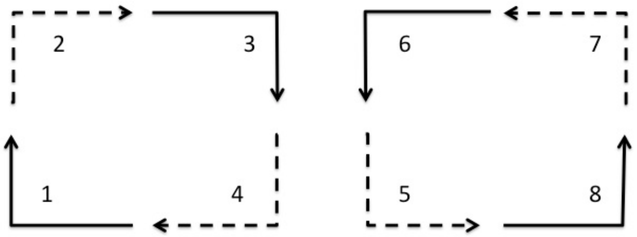


Fig. 5. Eight possible turns from a packet within a router.

B. Deadlock Problem

Due to the flexibility of the flexible router, any packet can be stored in any buffer regardless of its direction. This situation may lead to deadlock. For example, considering two neighbor routers connected through a channel in the X-axis (A and B), if all packets stored in router A are going to the Eastern direction and, at the same time, all packets in router B are going to the Western direction, a deadlock will occur, since there will be no available space in the destination buffers, and vice versa. This problem can be extended for more than two routers. The solution for this problem is analogue to the one used in [8]. As illustrated in Figure 5, there are eight possible turns that a packet can do inside a router. The turn-model [19] restricts the usage of two out of the eight turns. By restricting turns, the turn model imposes a total order on the allocation of the buffers, hence avoiding deadlocks.

C. Virtual Channel Allocation Method

Due to the aforementioned deadlock problem, a special Virtual Channel Arbitration algorithm must be implemented. This algorithm will implement the restrictions and will also give priority to allocate buffers in the vertical direction. Since we are using a X-Y algorithm, buffers of the X axis (West and East) might become heavily utilized. So, it makes sense to give priority to the allocation of the routers belonging to the Y axis (North or South).

The proposed mechanism consists of lending VCs from other input ports according to the availability of the buffers. If a packet wants to be transmitted and the input port is full, the FFC makes a request for an empty buffer in other ports. At every cycle, FFC look up the table trying to find an available buffer. If it fails, a request will be placed to another FFC. Even though the VC is available, it doesn't guarantee that it will be reserved to the requesting FFC; the same buffer can be requested for more than one FFC, in this case the owner of the buffer will choose a winner using a round-robin arbiter. In the buffer availability check, the FFC request and reply can be performed within one cycle, which means performance overhead. Moreover, as the channel demands for each router can vary in time for one or multiple applications, the buffer allocation of the flexible router is dynamic at *runtime*.

IV. Experiment Configuration and Analysis

A. Simulation Platform

To evaluate the efficiency of the proposed flexible router against the baseline router, a cycle-based accurate simulator called TOPAZ [20] was employed. TOPAZ is implemented in C++, models the pipeline of the routers, and operates at the granularity of individual architectural components. Simulations were performed using 64 routers organized as a 8x8 MESH network. In the simulator, we varied the buffer size, packet size, and quantity of virtual channels per input port over different traffic loads. Table I summarize the simulation parameter evaluation.

Each router has five physical bi-directional channels (ports) including the local port. The simulator keeps injecting messages into the network until it reaches 50.000 messages. A simple DOR-XY routing is used for all of our simulations where packets are first routed in the X-dimension followed by the Y-dimension. A DOR-XY was used because the main focus is on highlighting the improvement in performance due to the flexible router's adaptability, rather than the routing algorithm's. Wormhole flow control is also employed to control the flow of packets alongside the network, complementing previous evaluations. The tested network traffic pattern was the Uniform Random, where a node injects messages into the network at regular intervals specified by the injection rate. Normal Random distribution, where each node has the same probability of being chosen as a destination for a packet, was used for the destination node selection.

B. Analysis of Results

Our simulation exploration starts with the throughput comparison between the conventional, statically assigned buffer architecture in the base router compared to the proposed flexible router implementation. We started our discussions assuming that both use the same number of virtual channels. Results are shown varying the packet size from four flits up to sixteen flits for both of them. The injection load measures how many flits are injected in the network per cycle per router (flits/cycle/router), and the throughput is the amount of flits that leave the network per cycle (flits/cycle). Each line is labeled after the architecture in use and the packet size. For example, Flex 4 means that the line shows the performance of the scenario using a packet size of 4 and the flexible router architecture.

As the injection rate grows, the difference in throughput between the flexible and the base router increases. From the figures 6 to 8, we can see the difference for several buffer sizes using two virtual channels. It can be noticed that the difference in throughput grows up to 21% (4 flits per buffer, Figure 6) and the flexible router outperforms the base router in most cases 9% (8 flits per buffer, Figure 7) and 11% (16 flits per buffer, Figure 8). Another important result is that flexible saturates for higher injection loads when compared to the base one. The difference between both approaches is reduced as more resources are introduced to the network (larger buffers and increased amount of virtual channels). This is expected since the flexible router more efficiently manages routers with more limited number of resources. As the number of virtual channels grows, the flexible router starts to show a closer performance to the base router's. For four virtual-channels, the difference in performance between the two routers is small. Using packet size of 16 flits, flexible router outperforms the base router in 6% for buffer size of 4 and 8 flits (Figure 9), and 3% for buffer size of 16 flits (Figure 10 and Figure 11). Even though the flexible router outperform the base router in most cases (for packets of 4 flits and buffer size of 8 and 16 flits), the baseline router has respectively 10% and 8% better performance (Figure 10 and Figure 11).

As we can see in the charts, the best results are the ones in which the packet size matches exactly the size of the buffer size. An explanation for this is that packets can be moved entirely through VCs. By doing so, a packet doesn't interfere with other packets in the same buffer since no packets are allocated together. As the size of the packets grows, one needs more than one buffer to allocate these packets, in the worst case needing to allocate four buffers in different routers when the packet has sixteen flits and buffer size is four. The lack of resources increases the congestion of the network.

Another observation made in this paper is the performance of the flexible router using two virtual channels when compared with the performance of the baseline router using four virtual channels and both using four flits per buffer in an 8x8 mesh. As we can see in Figure 12, even though the baseline router has doubled the number of buffers, the throughput of the flexible router is only 3% inferior. When we added more two buffers in the baseline router, the contention problem was reduced improving its overall performance, but the flexible router was also able to overcome the same level of contention using just half the number of buffers. This implies great reductions in router area. As mentioned before buffers occupy most of the area of the router. This reduction is not only in the area but also in energy since the buffers are power-hungry components. A detailed discussion about the energy and the area savings caused by a flexible router can be found in [21].

Due to the inherent capacity of the flexible router to allocate resources in a more efficient way, the average bandwidth of the network is increased, which also increases the outflow of packets, and by doing so, mitigating the congestion problem.

V. Conclusion and Future work

In this paper, a perform analysis was done on the effects of the usage of virtual channels of flexible routers, as much as the employment of a wormhole flow control, which improved the

TABLE I
SIMULATION PARAMETER VARIATION

Parameter	Values
Packet Size	4, 8, 12, and 16
Buffer Size	4, 8, and 16
# of Virtual Channels	2 and 4

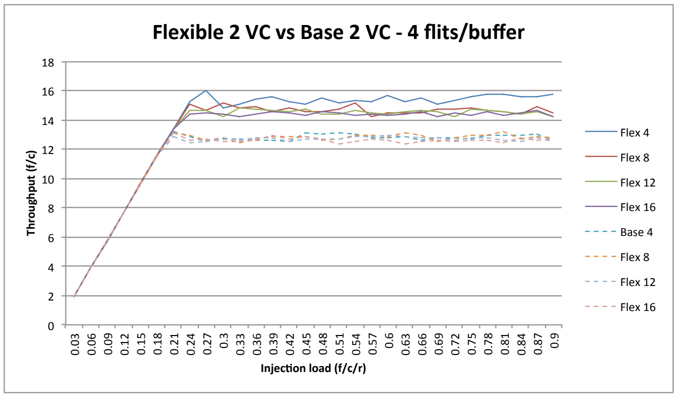


Fig. 6. Throughput - 2 Virtual Channels and 4 flits per buffer.

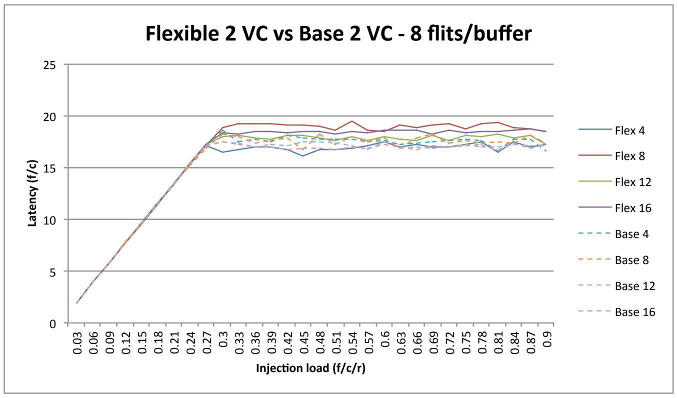


Fig. 7. Throughput - 2 Virtual Channels and 8 flits per buffer.

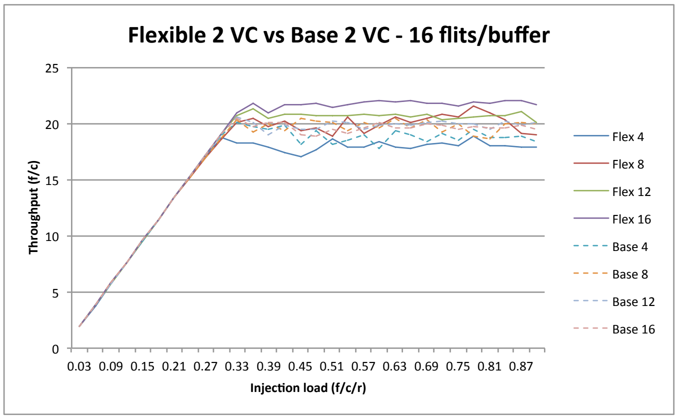


Fig. 8. Throughput - 2 Virtual Channels and 16 flits per buffer.

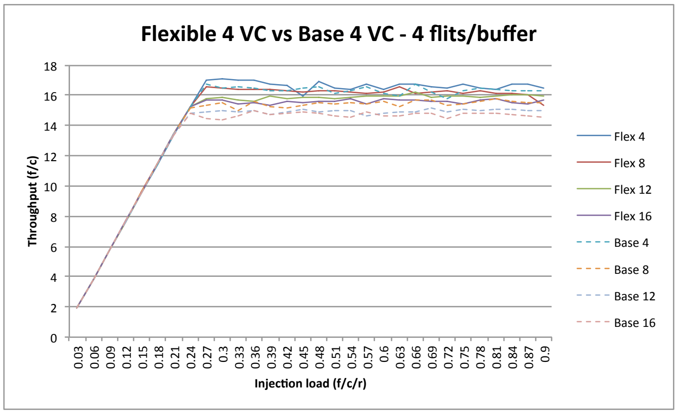


Fig. 9. Throughput - 4 Virtual Channels and 4 flits per buffer.

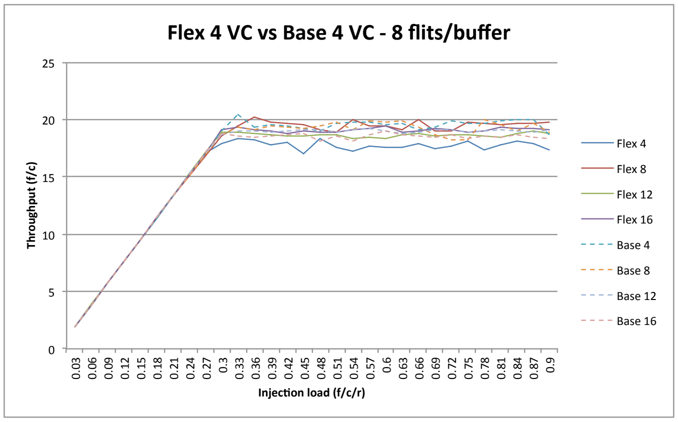


Fig. 10. Throughput - 4 Virtual Channels and 8 flits per buffer.

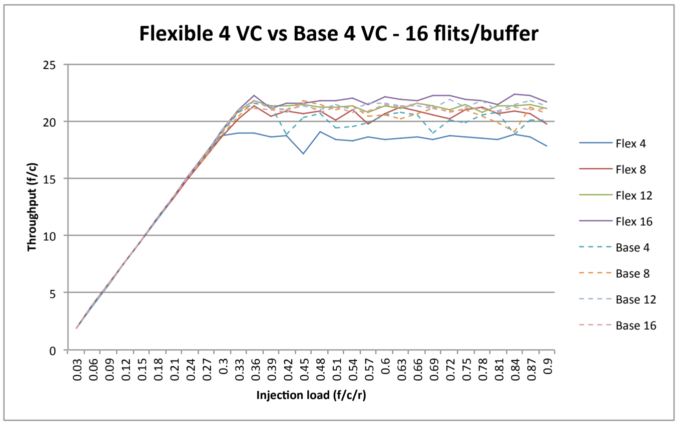


Fig. 11. Throughput - 4 Virtual Channels and 16 flits per buffer.

utilization of buffers proposed by the original flexible router and consequently enhanced its throughput and latency. Due to the before mentioned improvements, the contention problem was smoothed and the performance was increased up to 21% when buffers are small.

A flexible router was shown to be more efficient when the resources of the network were limited and had similar performance of the base router with double resources. Having presented the achievable improvements done by the flexible router, the new router architecture should be considered as an alternative for the base router when the network lacks resources when under heavy injection loads.

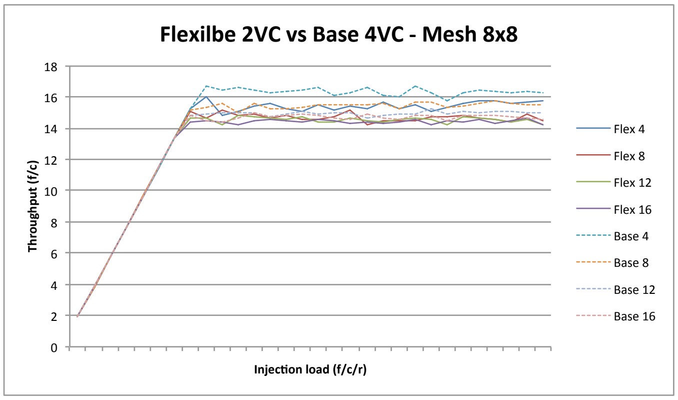


Fig. 12. Throughput - 2 Virtual Channels flexible and 4 Virtual Channels base.

As future work, we intend to test new routing approaches, traffic-patterns and workloads and traces from existing System-on-Chip architectures.

Acknowledgment

The authors would like to thank Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), Brazil, and Japan Student Services Organization (JASSO) Foundation for funding Israel M. Santos' work.

References

[1] L. Benini and G. D. Micheli, "Networks on Chips: A New SoC Paradigm," IEEE Computer, vol. 35, pp. 70-79, 2002.

[2] A. Hemani, A. Jantsch, S. Kumar, A. Postula, J. Oberg, M. Milberg, and D. Lindqvist, "Network on chip: An architecture for billion transistor era," in Proceedings of the IEEE NorChip Conference, pp. 65-79, 2000.

[3] W. J. Dally and B. Towles, "Route Packets, Not Wires: On-Chip Interconnection Networks," in Proceedings of the Design Automation Conference (DAC), pp. 684-689, 2001.

[4] C. Xuning and L. S. Peh, "Leakage power modeling and optimization in interconnection networks", in Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED), pp. 90-95, 2003.

[5] W. J. Dally and B. Towles, "Principles and practices of interconnection networks", Morgan Kaufmann, 2004 .

[6] T. T. Ye, L. Benini, and G. D. Micheli, "Analysis of power consumption on switch fabrics in network routers," in Proceedings of the 39th Design Automation Conference (DAC), pp. 524-529, 2002.

[7] Duato, J. , A new theory of deadlock-free adaptive routing in wormhole networks, IEEE Transactions on Parallel Distributed Systems, v. 4, n. 12, pp. 13201331, 1993.

[8] M. S. Sayed, A. Shalaby, M. E.-Sayed, and V. Goulart, "Flexible router architecture for network-on-chip". Computers & Mathematics with Applications, pp. 13011310, 2012.

[9] M. S. Sayed, A. Shalaby, M. Ragab, M. E.-Sayed, and V. Goulart, "Congestion mitigation using flexible router architecture for Network-on-Chip", Electronics, Communications and Computers (JEC-ECC), 2012 Japan-Egypt Conference on, pp.182-187, March, 2012.

[10] W. J. Dally and C. L. Seitz, "The torus routing chip,"Journal of Distributed Computing, vol 1(3), pp. 187-196, 1986.

[11] W. J. Dally, "Virtual-Channel flow control", in Proceedings of the 17th Annual Internation Symposium on Computer Architecture (ISCA), pp. 60-68, 1990.

[12] L. Benini and and G. De Micheli, "Networks on Chips: Technology and Tools", Morgan Kaufmann, 2006.

[13] M. J. Karol, M. G. Hluchyj, and S. P. Morgan, Input versus Output queuing on a space-division packet switch, IEEE Trans. Communication., Vol. 35, no. 12, pp. 1347-1356, December, 1987.

[14] R.S. Ramanujam, V. Soteriou, B. Lin , and Li-S. Peh, "Design of a High-Throughput Distributed Shared-Buffer NoC Router", Networks-on-Chip (NOCS), pp. 69-78, May, 2010.

[15] R. S. Ramanujam, V. Soteriou, B. Lin , and Li-S. Peh, "Extending the Effective Throughput of NoCs With Distributed Shared-Buffer Routers", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, v.30(4), pp. 548-561, April, 2011.

[16] C. A. Nicopoulos, D.Park, J. Kin, N. Vijaykrishnan, M. S. Yousif, R. Das Chita, "ViChaR: A Dynamic Virtual Channel Regulator for Network-on-Chip Routers". 2006 39th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO06), pp. 333346, December, 2006.

[17] Y. Tamir and G.L. Frazier, "High-performance multiqueue buffers for VLSI communication switches", in: Proceeding of the 15th Annual International Symposium on Computer Architecture, ISCA, pp. 343354, May, 1988.

[18] D. Matos, C. Concatto, F. Kastensmidt, L. Carro, A. Susin, and M. Kreutz, "Reconfigurable Routers for Low Power and High Performance", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, pp. 2045-2057, September, 2010.

[19] C.J. Glass, and L.M. Ni, "The turn model for adaptive routing", in: Proceeding of the 19th Annual International Symposium on Computer Architecture, ISCA, pp. 278287, May, 1992.

[20] P.Abad, P.Prieto, L.Menezo, A.Colaso, V.Puente, and J.A. Gregorio, "TOPAZ: An Open-Source Interconnection Network Simulator for Chip Multiprocessors and Supercomputers", Networks on Chip (NoCS), 2012 Sixth IEEE/ACM International Symposium on, pp. 99-106, May, 2012.

[21] H. El-Sayed, M. Ragab, M. S. Sayed., and V. Goulart, "Hardware Implementation and Evaluation of Flexible Router Architecture for NoCs", in Proc. of 20th IEEE Intl. Conf. on Electronics, Circuits and Systems, pp. 621-624, December, 2013.

Inferring TCP Congestion Control Algorithms by Correlating Congestion Window Sizes and their Differences

Toshihiko Kato, Atsushi Oda, Shun Ayukawa, Celimuge Wu, Satoshi Ohzahata
Graduate School of Information Systems
University of Electro-Communications
Chofu-shi, Tokyo, Japan
e-mail: kato@is.uec.ac.jp, oda@net.is.uec.ac.jp, s.aykw@net is.uec.ac.jp, clmg@is.uec.ac.jp, ohzahata@is.uec.ac.jp

Abstract— **Recently, according to the diversification of network environments, a lot of TCP congestion control mechanisms have been introduced. They define how a TCP sender increases the congestion window (*cwnd*) during no congestion and decreases *cwnd* when it detects congestion. Since the congestion control algorithms affect the performance of the Internet, it is important to know which algorithms are used widely. This paper proposes a scheme to infer the algorithm for individual TCP flows by examining the packet trace passively captured for the flows. Our scheme estimates *cwnd* values in RTT intervals and correlates the *cwnd* values and the differences of consecutive *cwnd* values. Our scheme aims to infer many of recently proposed congestion control algorithms, which have been out of scope in the conventional passive approaches. Our scheme adopts a simple approach just correlating *cwnds* and their differences, in contrast with the conventional approaches which estimate the internal TCP behaviors based on packet traces. This paper describes the details of our scheme and shows the results where our scheme is applied to an iPhone TCP communication.**

Keywords- TCP congestion control algorithms; passive monitoring; congestion window .

I. INTRODUCTION

Since the congestion control mechanism came to be used in Transmission Control Protocol (TCP) [1], only a few algorithms, such as Tahoe, Reno and NewReno [2], were used commonly for a long time. In the congestion control, a TCP sender transmits data segments under the limitation of the congestion window (*cwnd*) maintained within the sender, beside the advertised window reported from a TCP receiver. The value of *cwnd* increases as a sender receives ACK segments and is decreased when it detects congestions. How to increase and decrease *cwnd* is the key of congestion control algorithm. The early-stage algorithms mentioned above are summarized as an additive increase and multiplicative decrease (AIMD) because *cwnd* increases linearly and is reduced in an exponential fashion.

According to the diversification of network environments, many TCP congestion control algorithms have emerged [3]. For example, High Speed (HS) TCP [4], and CUBIC TCP [5] are designed for high speed and long delay networks. On the other hand, TCP Westwood [6] and its descendants are designed for lossy wireless links. While the algorithms mentioned so far are based on the packet losses, TCP Vegas [7] and FAST TCP [8] trigger congestion control against an increase of round-trip time (RTT). TCP Veno [9] and TCP Illinois [10] combine loss based and delay based approaches such that congestion control is triggered by packet losses but the delay determines how to increase *cwnd*.

The TCP congestion control algorithms affect the performance of the Internet, and so it is important to know which algorithms are used widely. Since the congestion control algorithm is implemented within a TCP sender, it cannot be identified from observable parameters in TCP segments. Instead, a tester which infers the algorithm needs to estimates internal behaviors of TCP senders from their input/output interactions.

The approaches to infer the congestion control algorithm are categorized into two groups. One is the passive approach where passively collected packet traces are examined to measure TCP behaviors. This approach has some limitations in the testing ability, but is non-intrusive and requires no additional equipment for measurement. The other is the active approach in which an active tester sends test inputs to a target node and checks the replies. This approach can perform a more comprehensive test than the passive one, but is limited to the case where a tester communicates with a node to be tested.

So far, several studies are proposed for both approaches [11]-[16]. However, as for the passive approach, there are no proposals on inferring the recently introduced algorithms. In this paper, we propose a new scheme based on the passive approach. The proposed scheme aims to infer many of recent congestion control algorithms and adopts a simpler methodology than the conventional studies.

The rest of this paper consists of the following sections. Section 2 surveys the related works specifically. Section 3 proposes our scheme. Section 4 gives some examples where our scheme is applied to an iPhone TCP communication. Section 5 gives the conclusions of this paper.

II. RELATED WORKS

As for the passive approach, TCPanaly [11] is one of the early stage research activities. It analyzes packet traces and tries to decide which implementation of TCP best matches the connection being observed.

Jaiswel et al. [12] adopted a similar approach with TCPanaly and proposes the TCP flavor identification among Tahoe, Reno and NewReno. Its basic idea is to construct three kinds of "replicas" of the TCP sender's state machine for individual TCP connections observed at the measurement

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-368-1

point. These replicas are for Tahoe, Reno and NewReno. For a segment sent by a TCP sender, each replica checks whether the segment is allowed or not. The numbers of violations are maintained and the TCP flavor with minimum number of violations is selected for the connection.

Oshio et al. [13] estimates the changes of *cwnd* values and extracts features, such as ratio of *cwnd* increase being one and so on. Based on these features, it discriminates one of two different versions randomly selected from thirteen TCP versions implemented in the Linux operating system.

Qian et al. [14], on the other hand, focuses on the extraction of statistical features based on the monitoring of one direction of TCP communications. They focused on the size of initial congestion window, the relationship between the retransmission rate and the time required to transfer a fixed size of data, which is used for detecting the irregular retransmissions, and the extraction of flow clock to find the TCP data transmission controlled by the application or link layer factors.

As an example of the active approach, TBIT [15] was developed to characterize the TCP behavior of major web servers. It checks the initial window size by not acknowledging any data segments sent by the server at the first data transfer. It also detects the congestion control algorithm by dropping two data segments (not acknowledging them) within one window. This discriminates Tahoe, Reno and NewReno.

CAAI [16] proposes the scheme to actively identify the TCP algorithm of a remote web server. It can identify all default TCP algorithms, such as AIMD and CUBIC, and most non-default TCP algorithms of major operating system families. It makes a web server send 512 data segments under the controlled network environment with specific RTT and observes the number of data segments contiguously transmitted without receiving any ACK segments. It then estimates the window growth function and the decrease coefficient, and using those estimations, determines the TCP algorithm for an individual web server.

As described above, CAAI proposes the inference of TCP congestion control algorithms used widely today, but no studies from the standpoint of passive approach. This paper proposes a passive monitoring based approach for inferring many of the TCP versions available today.

III. Proposal of Our Scheme

A. Design principles

The TCP congestion control algorithms have two parts. One is a part where a TCP sender increases *cwnd* at receiving an ACK segment acknowledging new data segments. The other is a part where a TCP sender decreases *cwnd* when it detects network congestion through retransmitting any data segments or perceiving an increase of RTT.

Our scheme to infer the congestion control algorithm is designed based on the following principles.

- Our scheme focuses on the increasing part of *cwnd*.
- It uses changes of the values of *cwnd* at individual RTT intervals.

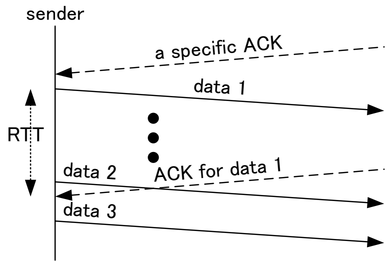


Figure 1. Principle for *cwnd* estimation.

- It estimates the value of *cwnd* at a moment when a TCP sender receives a specific ACK segment as the total size of inflight data segments, which are sent but not acknowledged, just before the TCP sender receives the ACK segment one RTT later than the first ACK. Fig. 1 shows this mechanism. Fig. 1 supposes that a sender receives *a specific ACK* and then sends *data 1*. After one RTT, the sender receives *ACK for data 1*. *Data 2* is the data segment sent out just before the sender receiving *ACK for data 1* and *data 3* is the data segment sent out just after the acknowledgment. Here, our scheme estimates the value of *cwnd* when the sender receives *a specific ACK* as
$$seqence\ number\ of\ data\ 2 + its\ length - ACK\ number\ of\ a\ specific\ ACK,$$ that is, $$seqence\ number\ of\ data\ 3 - ACK\ number\ of\ a\ specific\ ACK.$$
- The packet trace used in the inference may be captured in the middle of network. Therefore, in general, the packet sequence in the trace is different from the sequence in which the relevant TCP sender sends and receives packets. Our scheme needs to estimate the packet sequence in the TCP sender from that in the packet trace. For this purpose, our scheme utilizes the TCP time stamp option in TCP segments.
- Our scheme estimates a sequence of *cwnd* values observed in every RTT interval. We denote this sequence as $\{cwnd_i\}$. Then, a sequence of differences of consecutive *cwnd* values, $\{\Delta cwnd_i\}$, is defined by (1).
$$\Delta cwnd_i = cwnd_{i+1} - cwnd_i \qquad (1)$$
- In the end, our scheme evaluates the correlation of the two sequences, $\{cwnd_i\}$ and $\{\Delta cwnd_i\}$, by plotting them. The graphs depend on the congestion control algorithms.

It should be noted that, since our scheme does not require any tracking of TCP internal status, it is possible to infer the congestion control algorithms more easily than the conventional proposals.

The reason we adopt *cwnd* values at RTT intervals is as follows. First of all, many congestion control algorithms, such as Vegas and HS TCP, define a procedure for increasing *cwnd* at a RTT interval. Some algorithms, such as AIMD, specify a procedure for receiving individual ACK segments, but the purpose of those algorithms is the change of *cwnd* in a RTT interval. So, focusing the *cwnd* values at

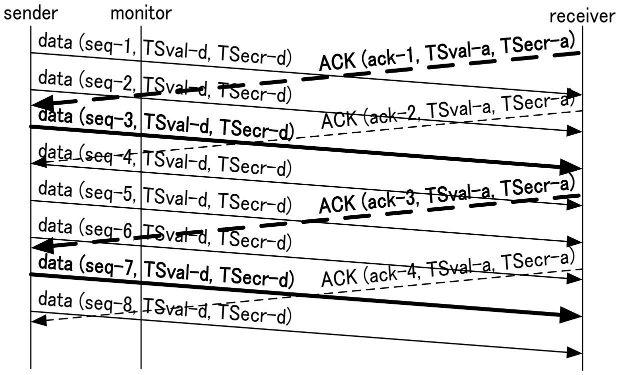

sender
monitor
receiver
data (seq-1, TSval-d, TSecr-d)
ACK (ack-1, TSval-a, TSecr-a)
data (seq-2, TSval-d, TSecr-d)
data (seq-3, TSval-d, TSecr-d)
ACK (ack-2, TSval-a, TSecr-a)
data (seq-4, TSval-d, TSecr-d)
data (seq-5, TSval-d, TSecr-d)
ACK (ack-3, TSval-a, TSecr-a)
data (seq-6, TSval-d, TSecr-d)
data (seq-7, TSval-d, TSecr-d)
ACK (ack-4, TSval-a, TSecr-a)
data (seq-8, TSval-d, TSecr-d)

Figure 2. Estimation of *cwnd* associated with one RTT.

RTT intervals is considered as appropriate for reflecting the purpose of congestion control algorithms.

The next point is that the algorithms define *cwnd* values in a byte but data segments are sent in the unit of maximum segment size (MSS). Therefore, the passive approach can detect the change of *cwnd* values in the order of MSS. On the other hand, many algorithms change *cwnd* values in the order of MSS during a RTT interval.

B. Methodology for estimating cwnd at RTT intervals

As mentioned above, the time associated with individual captured segments in a packet trace is not the exact time when the data sender sent or received those segments. So, our scheme estimates a *cwnd* associated with one RTT interval in the following way.

- ➢ First, our scheme focuses on an ACK segment in the packet trace, for example, *ACK with ack-1* in Fig. 2.
- ➢ Next, it looks for the first data segment whose TSecr (Time Stamp Echo Reply) is equal to TSval (Time Stamp Value) of the ACK segment we are focusing on, *data with seq-3* in Fig. 2.
- ➢ Our scheme then looks for the first ACK segment acknowledging this data segment, in this case, *ACK with ack-3*.
- ➢ As the fourth step, it looks for the data segment whose TSecr is equal to TSval of the second ACK segment. In the example of Fig. 2, this corresponds to *data with seq-7*.

After these steps, our scheme estimates that the first data segment, *data with seq-3*, and the second ACK segment, *ACK with ack-3*, construct a RTT relationship. Based on this estimation, our scheme estimates that *cwnd* in the unit of MSS at the moment of receiving *ACK with ack-1* is equal to (2).

$$\frac{seq\text{-}7 - ack\text{-}1}{MSS} \tag{2}$$

C. Applying our scheme to AIMD

In the AIMD congestion control algorithm, *cwnd* is increased each time the TCP sender receives an ACK segment acknowledging new data. The increase is one segment during the slow start phase, and $\frac{1}{cwnd}$ segments during the congestion avoidance phase. During one RTT, *cwnd* of data segments are sent and acknowledged. Therefore, in the slow start phase, *cwnd* is increased by the value of *cwnd* (*cwnd* is doubled). This means that

$$\Delta cwnd_i = cwnd_i.$$

On the other hand, in the congestion avoidance phase, *cwnd* is increased by one segment during one RTT. So, in this phase,

$$\Delta cwnd_i = 1.$$

So, plotting *cwnd* and $\Delta cwnd$ generates the graph in Fig. 3. In this figure, it is assumed that the initial congestion window is one MSS, and that the slow start continues until *cwnd* is 16 followed by the congestion avoidance.

D. Applying our scheme to TCP Vegas

TCP Vegas detects congestion by the increase of RTT. It measures the minimal RTT during the connection lifetime. With the current values of *cwnd* and RTT, it estimates the buffer size in the bottleneck node as (3).

$$BufferSize = cwnd \times \frac{RTT - RTT_{min}}{RTT} \tag{3}$$

Vegas uses this *BufferSize* for the control in the congestion avoidance phase in the following way.

- ● If $BufferSize < \alpha$, then *cwnd* is increased by one MSS. (In the Linux implementation, α is less than 2.)
- ● If $BufferSize > \beta$, then *cwnd* is decreased by one MSS. (In the Linux implementation, β is more than 4.)
- ● If $\alpha \le BufferSize \le \beta$, then the system is considered to be in a steady state and no modification to *cwnd* is applied.

This examination is done at every RTT interval. Therefore, the difference of *cwnd* at RTT interval, $\{\Delta cwnd_i\}$, is

$$\Delta cwnd_i = 1, 0\ or -1$$

in the congestion avoidance phase.

As for the slow start phase, *cwnd* is increased every other RTT. This means that $\{\Delta cwnd_i\}$ is

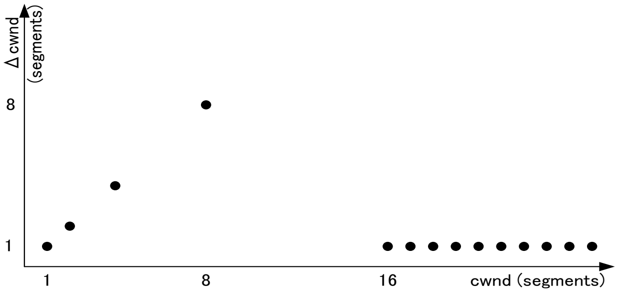


Figure 3. Applying to AIMD.

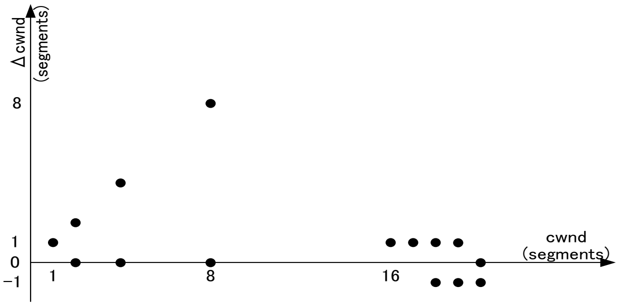


Figure 4. Applying to TCP Vegas.

$$\Delta cwnd_i = cwnd_i \ or\ 0$$

in this phase. So, plotting $cwnd$ and $\Delta cwnd$ generates the graph in Fig. 4 for TCP Vegas. In this figure, it is assumed that the initial congestion window is one MSS, that the slow start continues until *cwnd* is 16, and that *BufferSize* increases when *cwnd* is 20.

E. Applying our scheme to TCP Veno

The Veno (VEgas and ReNO) algorithm uses the Vegas estimate in order to limit the increase of *cwnd* during the congestion avoidance phase. If the Vegas buffer estimate shows excessive buffer utilization (i.e., $BufferSize > \beta$), a TCP sender increases *cwnd* by one for every two RTT.

This means that the increase of *cwnd* during the congestion avoidance phase is

$\Delta cwnd_i = 1$ during no congestion, and
$\Delta cwnd_i = 1\ or\ 0$ during congestion.

As a result, the plotting of *cwnd* and $\Delta cwnd$ will be as the graph in Fig. 5 for Veno. In this figure, it is assumed that the initial congestion window is one MSS, that the slow start continues until *cwnd* is 16, and that congestion occurs when *cwnd* is 20.

F. Applying our scheme to HS TCP

The HS TCP changes the increase coefficient α according to the current size of cwnd. Here, α defines how many segments are added to *cwnd* for one RTT in the congestion avoidance phase. When cwnd is less than or equal to 38 segments, α is 1, which has the same behavior as the traditional AIMD. If *cwnd* is more than 84K segments, α is 70. Between 38 and 84K segments, α is interpolated from 1 and 70 linearly.

In the slow start phase, HS TCP adopts the limited slow start, which bounds the maximum increase step during this phase to 100 segments.

These specifications give the plotting of *cwnd* and $\Delta cwnd$ as shown in Fig. 6 in the form of semilog graph. In the graph, the congestion avoidance is started from *cwnd* of 32, and the relationship between the consecutive *cwnds* is defined as in (4) for $cwnd_i$ which is between 38 and 8700.

$$cwnd_{i+1} = cwnd_i + \frac{70-1}{8700-38}(cwnd_i - 38) + 1 \qquad (4)$$

G. Applying our scheme to CUBIC TCP

CUBIC TCP defines cwnd as a cubic function of elapsed time T since the last congestion event. Specifically, it defines cwnd by (5).

$$cwnd = C\left(T - \sqrt[3]{\beta \cdot \frac{cwnd_{max}}{C}}\right)^3 + cwnd_{max} \qquad (5)$$

Here, C is a predefined constant, β is a coefficient of multiplicative decrease in the congestion control, and $cwnd_{max}$is the value of *cwnd* just before the loss detection in the last congestion event.

From this equation, the increase of *cwnd* during one RTT can be obtained approximately by (6).

$$RTT \cdot \frac{d(cwnd)}{dT} = RTT \cdot 3C\left(T - \sqrt[3]{\beta \cdot \frac{cwnd_{max}}{C}}\right)^2 \qquad (6)$$

By using (5), (6) is represented as a function of *cwnd* as in (7).

$$RTT \cdot \frac{d(cwnd)}{dT} = 3RTT \cdot \sqrt[3]{C}\left(\sqrt[3]{cwnd - cwnd_{max}}\right)^2 \quad (7)$$

This result gives the plotting of *cwnd* and $\Delta cwnd$ as in Fig. 7. Here, it is assumed that $cwnd_{max}$ is 0 in the slow start phase and $3RTT \cdot \sqrt[3]{C} = 1$.

H. Applying our scheme to TCP Illinois

TCP Illinois changes the increase coefficient of *cwnd*, α, according to the queuing delay. The queuing delay is measured as the increase of RTT from the minimum RTT for the connection. Depending on the queuing delay, α changes from 0.1 segments to 10 segments. The value of α is updated once per every RTT. Therefore, the plotting of *cwnd* and $\Delta cwnd$ will be given as in Fig. 8. In this figure, it is assumed that the initial congestion window is one MSS,

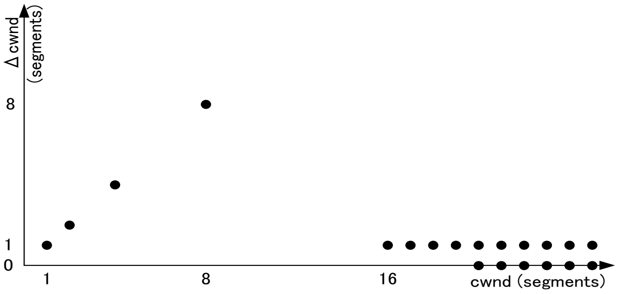


Figure 5. Applying to TCP Veno.

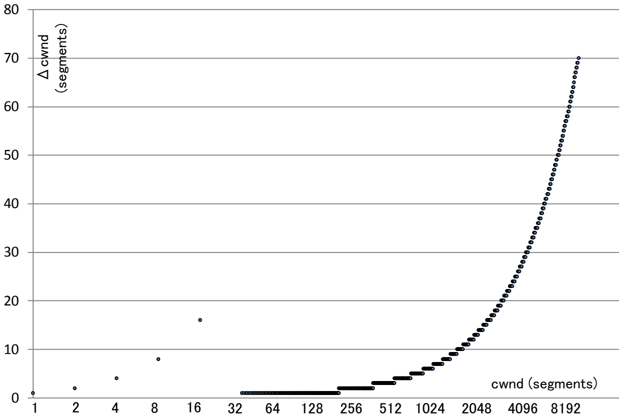


Figure 6. Applying to HS TCP.

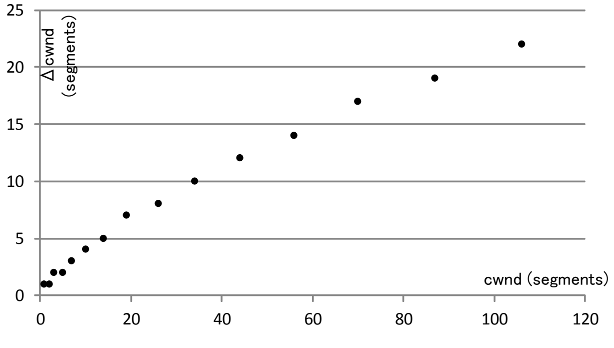


Figure 7. Applying to CUBIC TCP.

ISBN: 978-1-61208-368-1

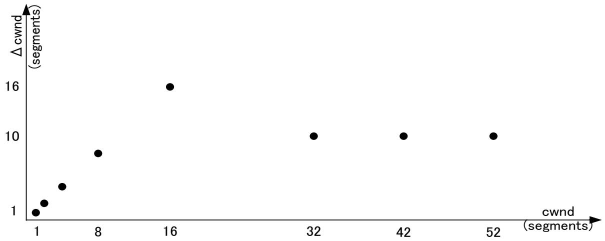


Figure 8. Applying to TCP Illinois.

that the slow start continues until *cwnd* is 16, and that the queuing delay is small in the beginning of the congestion avoidance.

IV. Inferring iPhone 5 TCP algorithm

As an example of the inferring of TCP congestion control algorithm using our scheme, we performed an experiment estimating the TCP algorithm of iPhone 5. Fig. 9 shows the configuration. An ftp application on an iPhone 5 terminal communicates with an ftp server through an LTE network and the Internet. While the iPhone 5 uploads a file to the server, it moves on a local train in Tokyo. The packet traces are collected in a PC connected with the iPhone through the remote virtual interface [17].

We collected two packet traces. Fig. 10 shows the results of the first example. Fig. 10 (a) and (b) show the TCP sequence number versus time and *cwnd* value versus time in this communication, respectively. These graph show that the handover happened two times around at 7 second and 35 second, and accordingly, the *cwnd* value decreases. The graph (c) shows the relationship between Δ*cwnd* and *cwnd*. It is noted that the decreases of *cwnd* are not described in this figure. The graph shows the slow start like behavior from *cwnd* =1 to *cwnd* = 23, in which Δ*cwnd* is proportional to *cwnd*. On the other hand, from *cwnd* = 30 to 111, most of observed values for Δ*cwnd* is equal to one. It can be said that the graph in Fig. 10 (c) is quite similar with that in Fig. 3 and so the results of this example says that the TCP congestion control algorithm used in iPhone 5 is AIMD.

Fig. 11 shows another example for iPhone 5. In this example, packet losses occur at 20, 30, 70 and 95 second in the communication, and accordingly the *cwnd* value changes as shown in Fig. 11 (b). Based on this graph, we depicted the relationship between Δ*cwnd* and *cwnd* as shown in Fig. 11 (c). This figure has a proportional part and one segment part similarly with Fig. 10 (c). This result also says that the the TCP congestion control algorithm used in iPhone 5 is AIMD.

V. Conclusions

This paper proposed a simple but effective scheme inferring the TCP congestion control algorithm from passively collected packet traces. The proposed scheme estimates cwnd values at every RTT intervals from packet traces and makes the correlation beween the *cwnd* values and the differences consecutive *cwnd* values by plotting these values. We showed that the result plotting can explicitly

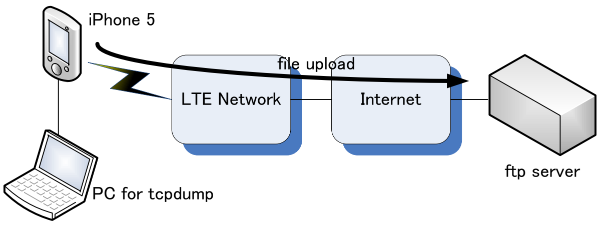
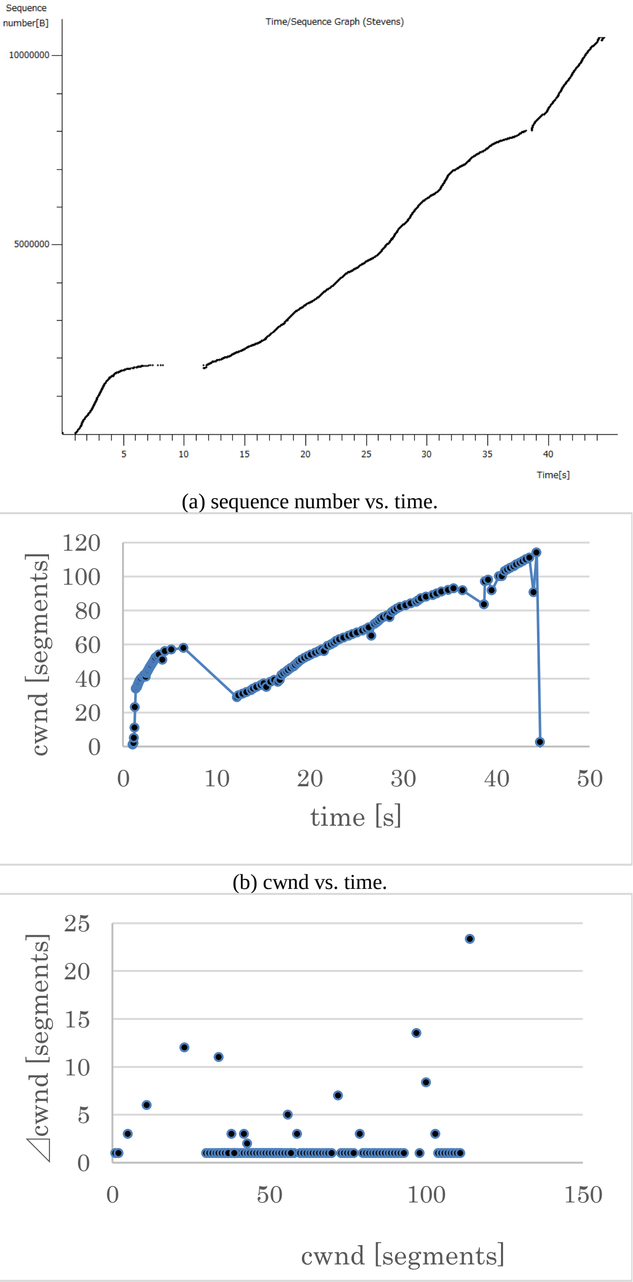


Figure 9. Configuration of experiment.



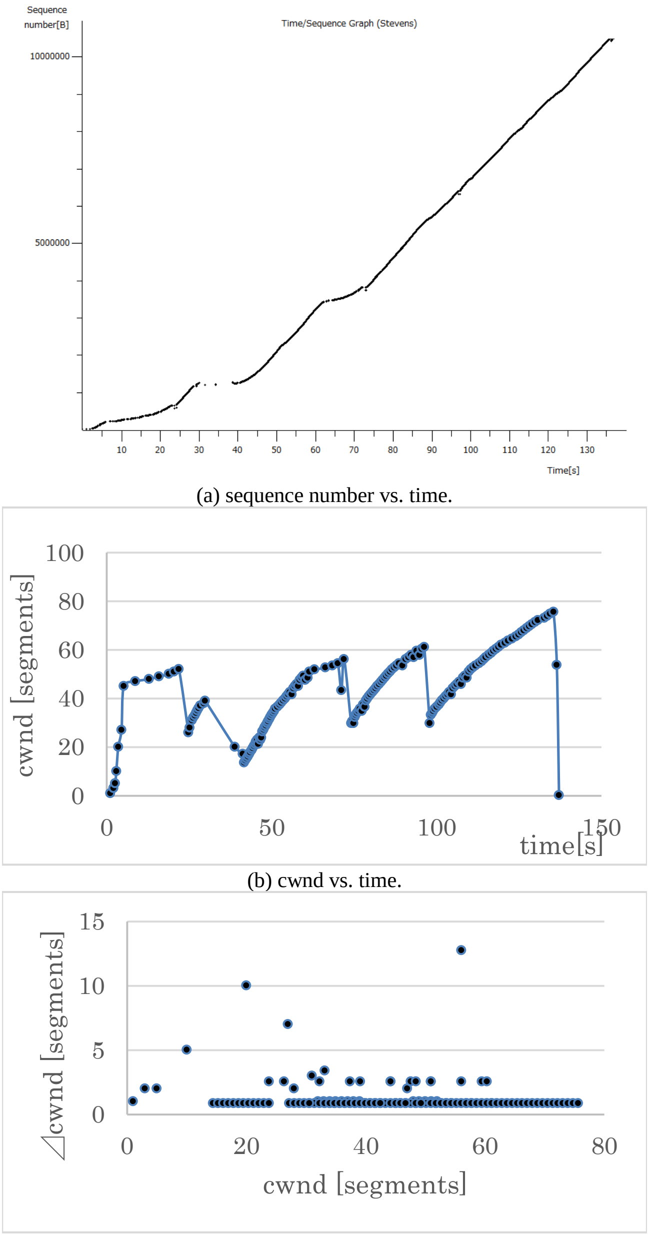
(a) sequence number vs. time.

(b) cwnd vs. time.

(c) Δcwnd vs. cwnd.

Figure 10. Results of first example.

distinguish AIMD, TCP Vegas, TCP Veno, HS TCP, CUBIC TCP and TCP Illinois. As an example, we applied our scheme to identify the TCP congestion control algorithm



(a) sequence number vs. time.

(b) cwnd vs. time.

(c) Δcwnd vs. cwnd.

Figure 11. Results of second example.

used in iPhone 5. From two packet traces in which an iPhone 5 terminal is sending ftp data, our scheme showed two graphs showing the AIMD like relationship between $\Delta cwnd$ and $cwnd$. We could successfully conclude that the algorithm used in iPhone 5 is AIMD from these results.

References

[1] V. Javobson, “Congestion Avoidance and Control,” ACM SIGCOMM Comp. Commun. Review, vol. 18, no. 4, Aug. 1988, pp. 314-329.

[2] S. Floyd, T. Henderson, and A. Gurtov, “The NewReno Modification to TCP’s Fast Recovery Algorithm,” IETF RFC 3728, April 2004.

[3] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, “Host-to-Host Congestion Control for TCP,” IEEE Commun. Surveys & Tutorials, vol. 12, no. 3, 2010, pp. 304-342.

[4] S. Floyd, “HighSpeed TCP for Large Congestion Windows,” IETF RFC 3649, Dec. 2003.

[5] S. Ha, I. Rhee, and L. Xu, “CUBIC: A New TCP-Friendly High-Speed TCP Variant,” ACM SIGOPS Operating Systems Review, vol. 42, no. 5, July 2008, pp. 64-74.

[6] S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, and R. Wang, “TCP Westwood: Bandwidth estimation for enhanced transport over wireless links,” Proc. ACM MobiCom ’01, July 2001, pp. 287-297.

[7] L. Brakmo and L. Perterson, “TCP Vegas: End to End Congestion Avoidance on a Global Internet,” IEEE J. Selected Areas in Commun., vol. 13, no. 8, Oct. 1995, pp. 1465-1480.

[8] D. Wei, C. Jin, S. Low, and S. Hegde, “FAST TCP: Motivation, Architecture, Algorithms, Performance,” IEEE/ACM Trans. on Networking, vol. 14, no. 6, Dec. 2006, pp. 1246-1259.

[9] C. Fu and S. Liew, “TCP Veno: TCP Enhancement for Transmission Over Wireless Access Networks,” IEEE J. Selected Areas in Commun., vol. 21, no. 2, Feb. 2003, pp. 216-228.

[10] S. Liu, T. Bassar, and R. Srikant, “TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks,” Proc. VALUETOOLS ’06, Oct. 2006.

[11] V. Paxson, “Automated Packet Trace Analysis of TCP Implementations,” ACM Comp. Commun. Review, vol. 27, no. 4, Oct. 1997, pp.167-179.

[12] S. Jaiswel, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley, “Inferring TCP Connection Characteristics Through Passive Measurements,” Proc. INFOCOM 2004, March 2004, 1582-1592.

[13] J. Oshio, S. Ata, and I. Oka, “Identification of Different TCP Versions Based on Cluster Analysis,” Proc. ICCCN 2009, Aug. 2009, pp. 1-6.

[14] F, Qian, A. Gerber, and Z. Mao, “TCP Revisited: A Fresh Look at TCP in the Wild,” Proc. IMC ’09, Nov. 2009, pp. 76-89.

[15] J.Padhye and S. Floyd, “On inferring TCP behavior,” Proc. ACM SIGCOMM, Aug. 2001, pp.287-298.

[16] P. Yang, W. Luo, L. Xu, J. Deogun, and Y. Lu, “TCP Congestion Avoidance Algorithm Identification,” Proc. ICDCS ’11, June 2011, pp. 310-321.

[17] Apple Inc. “Technical Q&A QA 1176 Getting a Tacket Trace,” Available from:

https://developer.apple.com/library/ios/qa/qa1176/_index.html#//apple_ref/doc/uid/DTS10001707, 2014.08.13.

The K-means and TSP Based Mobility Protocol Modeling as a Probabilistic Combinatorial Optimization Problem

Monia Bellalouna[1], Afef Ghabri[2], Walid Khaznaji[3]
[13]Laboratory CRISTAL POLE GRIFT, [2]Laboratory RIADI-GDL
University of Manouba, National School of Computer Sciences (ENSI), 2010
Manouba, Tunisia
e-mail: [1]monia.bellalouna@ensi.rnu.tn, [2]afef2108@hotmail.com, [3]mwkhaznaji@yahoo.fr

Abstract— **Fault tolerance is considered as a critical issue and a very interesting subject of research in Wireless Sensor Networks (WSN). Some sensors may be blocked or may fail due to a lack of energy or because of their manufacture. The external interactions (interferences, malicious attacks) can also be the source of malfunctions. The failure of sensors should not affect the network performance. This is a problem of reliability or fault tolerance which is the ability to maintain network functionality without interruptions due to a failure of a sensor node. It therefore aims to reduce the influence of these failures on the overall task of a wireless sensor network. Protocols and fault- tolerant approaches must be used to ensure reliable delivery of data packets to the base station and to guarantee reliable functioning even after the vulnerability of some network components. In this paper, we describe the K-means And Traveling Salesman Problem-based mobility protocol used to assure the proper functioning of the networks; we also propose a theoretical modeling of a probabilistic combinatorial optimization problem, which is explored through this method in order to minimize the energy consumption and improve fault tolerance for WSN.**

Keywords- Wireless sensor networks; failure; fault-tolerance; modeling; probabilistic.

I. INTRODUCTION

Wireless sensor networks represent a very promising domain that has become a new research focus in communication and computer fields. They can be used in different environments in a random way and in a large variety of applications due to their easy deployment and their low cost of construction. To achieve some applications, such as medical care, military surveillance, disaster relief and environmental monitoring, reliability is fundamental and essential. A wireless network is composed of plenty of sensors that are deployed in the monitoring field, and that have perception, processing and communication ability. However, the limitation of energy in wireless nodes, and hostile environments in which they could be used, are factors that make such networks very vulnerable [1][2]. In fact, they are subject to different forms of breakdowns that affect their reliability. These problems include computer attacks and hardware failures which are, nowadays, a real threat constantly growing. The failures of sensors can be caused by different reasons, including the physical damages, malicious attacks, environmental interferences, communication link errors or the depletion of energy [2]. Without successful transmission and secure routing, many applications of wireless sensor networks cannot work.

Because of their sensitivity, several research projects have been conducted in order to find solutions to these networks in the presence of failures and intrusions [3]. In fact, a sensor network must be able to maintain and keep its functionality without interruptions caused by the failures of nodes. In other words, these breakdowns should not affect the overall functioning of the network. This problem of fault tolerance has seen a great significance among various fields of research in wireless sensor networks. So, one of the basic challenges is to guarantee the safety and the proper functioning of equipments by developing fault-tolerant and robust algorithms that offer reliability of transmitted data. The purpose of this paper is to propose a possible direction of future researches by including a probabilistic modeling of the K-means And Traveling Salesman Problem-based mobility protocol [4], which is a fault tolerant protocol in wireless sensor networks. The rest of the paper will be organized as follows: In the second Section, we will present an overview of the fault tolerance problem. Section 3 will describe the principle of this relevant a priori protocol. A theoretical probabilistic modeling of it will be also proposed in Section 4. Section 5 presents the simulation results and analysis. Finally, we will conclude the paper in Section 6.

II. FAULT TOLERANCE IN WSN : PROBLEM PRESENTATION

The communications between the collector node and the other sensors of a network need the implementation of routing protocols that are based on multi-hop communication. Each sensor then acts as a router in addition to its role as a data source. However, faults can happen because of some problems such as the lack or the loss of energy, the interferences of the environment (heat, rain) or by a destructive agent (like animals), attacks (Sybil, Wormhole, Selective forwarding, sinkhole, etc.) and also the physical damages. In this case, it will be possible that one or many sensors do not operate. This causes the loss of communication links that leads to a significant change in the entire network topology. The fault is the primary source of an error that causes the system failure [5]. So, the network connectivity can be affected and its life will be decreased. In this case, it must be able to detect this error and to remedy it, by finding another way to transmit information and to maintain the network always operational.

ISBN: 978-1-61208-368-1

We can say that the goal of fault tolerance is to avoid the total flaw of the system despite the existence of errors in a subset of its elementary components. The tolerance degree depends on the nature of the application, its degree of criticality and on the exchanged data. Then, it is essential to provide fault tolerant protocols that allow us to choose the best paths in order to route information from the source to the collector. They also permit the selection of an alternative path if there was a failure while sending data on the initial route, in case of an interruption at one or several sensor nodes of it. In addition, the implementation of fault tolerant clustering protocols is useful in order to provide a better routing management [6]. Then, there are algorithms used to determine many routes from each node to the sink, which guarantees the presence of more than one reliable path for transmission. This provides a fast transfer resumption in case of failure on the first selected route (selecting one of the remaining routes). Other protocols realize a better management of energy use, with the aim of increasing the network lifetime [7]. Many works in the literature suggest fault tolerant methods in sensor networks, such as the Energy Aware Routing "EAR" protocol [8] and the Periodic, Event-driven, Query-based "PEQ" protocol [8], for many objectives. As each has been developed to achieve a specific purpose, they vary widely in many settings, including security, accuracy, configurability, cost and reliability. In this work, a probabilistic derivation of the reliability problem is considered.

Actually, although there are a large number of theoretical models for problems coming from the real world, the application of these models, in a direct way, is difficult and sometimes impossible due to the vagueness, the inaccuracy or the lack of data. In some contexts, based on estimations or statistical measures, a solution may be required even before the specification of information. In fact, there are several applications where obtaining current data in a certain way is not possible due to the volatile nature of data. The community of operational research has introduced several optimization frameworks to address these constraints [9]. In our work, a sensor network is modeled by a graph, and we are operating in a situation where the vertices do not exist in a deterministic way in this graph, but they are present in a probabilistic manner. In other words, a probability of presence will be associated with each vertex. This work is located in a study framework of combinatorial optimization problems when their proceedings are changing in a probabilistic way [10].

The aim is to present a priori strategy which ensures the reliability of data transmission in the presence of faults: on any instance of the problem, we avoid the total flaw of the system by changing the graph structure (transformation in a subgraph) according to a modification strategy that will be specified in advance. Fault tolerance in wireless sensor networks can then be presented as a probabilistic optimization problem that will be modeled based on different protocols and particularly the K-means And Traveling Salesman Problem-based mobility protocol, in which we will be interested in the following sections.

III. K-means And TSP-based mobility "KAT-mobility" Protocol

Nakayama et al. [12] proposed the K-means And TSP-based mobility "KAT-mobility" protocol based on optimization algorithms of routing and aggregation. It is a model of mobility of sinks that can effectively collect detected data used in a wireless sensor network, even if some sensors are destroyed [11]. The system is composed of two modules: the clustering algorithm and the approximate solution for the TSP "Traveling Salesman Problem" [12]. The sensors are firstly divided into groups by using the clustering algorithm, from which the centers of the groups are determined as anchor points. The route of the mobile sink is determined as an approximate solution of the TSP. Here, it is assumed that an administrator distributes sensors to supervise the targeted area, and the sensors are scattered at random positions and do not move afterwards. After the grouping of sensor nodes, this method reaches the mobile sink to make a course through the centers of groups according to the trajectory of an optimized path. The mobile sink collects then the data coming from sensors on the level of the visited groups. Its trajectory is assumed to be random in order to mitigate malicious attacks. The compromise between the flow rate and the energy consumption is regarded as the efficiency metric during the evaluation. During this time, the KAT-mobility protocol can calculate the migration route for the sink in order to get around the damaged area or malfunctioned sensors due to attacks while preserving its random behavior. In other words, after the reorganization of the network into groups, the proposed approach pilots the mobile collector to move through the centers of the groups by taking the optimal route. Thus, the mobile collector recovers the information from the sensors of the visited groups. The principle of this protocol is summarized in these two procedures: the optimization of the routing path and the clustering. Figure 1 shows this principle [12].

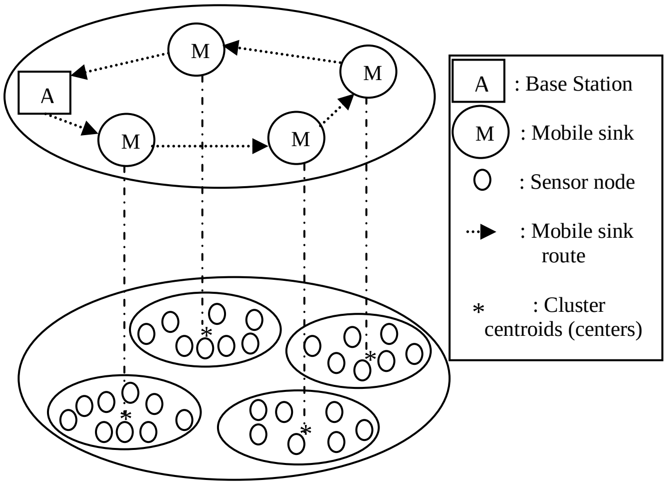


Figure 1. KAT-mobility protocol

A. Clustering Procedure

A sensor network is often composed of several thousands sensor nodes. To reduce the complexity of the routing algorithms, to facilitate data aggregation, to simplify the network management and to optimize energy consumption, sensors are grouped into clusters. The nodes that are grouped together in a cluster will be easily able to communicate with each other. A cluster head is elected to carry out several tasks, such as filtering, fusion and aggregation, with the possibility to be changed if it fails or if it reaches its power limit [13]. All communications of all nodes will be made through the head of the cluster to which they belong. The KAT-mobility protocol is based on clustering and especially on the K-means method [14][15], in which the cost of a group is estimated by the approximation error between the nodes and the collector. This algorithm divides the set of sensors virtually into K clusters (C_1, C_2 … C_K) geographically close. We denote by N the number of nodes in the network, usually N >> K. Let m_j (j = 1, 2 ... K) be a collector and x_i (i = 1, 2 ... N) be a sensor node, which is represented by a 2-dimensional vector (i.e., position of the node). d (x_i, m_j), which is indicated by the Euclidian distance between the collector (group center) and the sensor, represents the approximation error. The goal is to assign each node to a cluster C_j by reducing the total error of the clusters in order to reduce energy consumption [16]. The sum of approximate errors is expressed in the following formula:

$$DT\big(C_1(m_1), C_2(m_2), ..., C_K(m_K)\big) = \sum_{i=1}^{K} \sum_{x_i \in C_j} d(x_i, m_j) \quad (1)$$

The goal is to minimize the energy of communications by clustering in order to reduce the battery consumption of each sensor, which is proportional to d (x_i, m_j) [17]. The final objective is to assure the configuration of C_j such that DT is minimized. The clustering module of the KAT-mobility protocol can be summarized as follows [4]:

1. Initialize the location m_j randomly, t=0.

2. Define the threshold "THR" which is the stopping criterion of the following iterative process.

    - When $d^{(t)}$ (x_i, m_j) < $d^{(t)}$ (x_i, m_{j*}), ∀ j ≠ j*, assign a node $x_i \in C_j^{(t)}$.

    - Set the collector positions at the center of each group.

    $$m_j^{(t+1)} = \frac{1}{\left|C_j^{(t)}\right|} \sum_{x_i \in C_j^{(t)}} x_i \quad (2)$$

    - Calculate the sum of approximation errors $DT^{(t+1)}$ at time (t+1), and if

    $$\frac{\left|DT^{(t+1)} - DT^{(t)}\right|}{DT^{(t)}} \rangle \; THR \quad (3)$$

    is true, t can be updated and the iteration will be continued. Otherwise, the iterative process is stopped and the final center is set to $m_j^{(t+1)}$.

After the realization of the clustering procedure, the second module of the KAT-mobility protocol begins.

B. Routing Path Optimization

Finding the best route for the mobile node is analogous to the TSP. A sink represents then the traveling salesman and the cluster centroids define cities. The mobile sink passes through the clusters and gathers data coming from various nodes. As it is possible to increase efficiency by reducing the travel time, it is preferable that the sink traces the shortest route through the cluster centroids. The path optimization of the mobile collector to visit once and only once every cluster centroid is equivalent to searching for the shortest trip of the traveling salesman in order to visit each city once [4]. However, this problem is NP-hard [18]; therefore, the conclusion of the optimal trajectory can not be realized in an easy way. In fact, it is one of the most studied combinatorial optimization problems that its difficulty reveals especially through the large number of solutions. In order to solve this problem, a particular family of algorithms called the heuristics, ensuring the obtaining of almost optimal solutions, is proposed. For the KAT-mobility protocol, the local search algorithms Or-Opt and 2-Opt, based on the modification of a current solution to TSP by heuristics, are implemented [19][20]. Fortunately, for wireless sensor networks, nodes can communicate together, and the mobile sink does not need to visit all the nodes. After clustering the nodes, we only need to optimize the paths among cluster centroids. The mobile sink traces then the trajectory of the optimal TSP solution. Using the same formula mentioned in [21], the objective is finding the Hamiltonian path π that reduces the tour lengths.

$$\sum_{j=0}^{k-1} d\Big(m_{\Pi(j)}, m_{\Pi(j+1)}\Big) + d\Big(m_{\Pi(k)}, m_{\Pi(0)}\Big) \quad (4)$$

where the initial location of the mobile sink is m_0. This quantity mentioned in (4) denotes the tour length of a sink that will be carried out by visiting the centers in the specified order according to the permutation and while returning to the starting position [21]. The KAT-mobility method assumes that the mobile collector has a priori knowledge of the locations of its member sensors [4]. It is possible that the collector lose communications with its blocked or faulty member nodes. In this case, it can stay at the center of its cluster to discover broken nodes and its trajectory can then be recalculated as soon as it reaches the access point. Updating this path is preceded by a modification of the centers positions in the network, i.e., a new clustering procedure through the subset of functional nodes is performed [4]. Consequently, it will be more interesting to model the KAT-mobility method as a probabilistic combinatorial optimization problem.

IV. Probabilistic Modeling

We model the wireless network by a graph G (V, E), where V is the set of cluster centroids and E ⊆ V^2 represents the set of edges reflecting the possible communications between these points. The pair (m_1, m_2) belongs to E if and only if m_2 is the neighbor of m1. We denote by K the number of vertices in the graph (|V| = K), which is considering here as the problem size. In wireless sensor networks, one or several sensors may not function correctly and in order to avoid the total flaw of the system despite the presence of faults in a subset of its elementary components, two approaches can be used: the re-optimization strategy treated by Nakayama et al. [4] and the a priori strategy that represents our proposal through this paper.

A. Re-optimization Strategy

Frequently in applications, after having solved a particular exemplary of a given combinatorial optimization problem, we must solve repeatedly many copies of the same problem. These additional copies are generally simple variations of the original problem; however, they are sufficiently different to require an individual treatment [8]. The most natural approach used to address this kind of situation consists in solving in an optimal way the different potential copies. We call this strategy "the re-optimization strategy" [22]. However, it has many disadvantages and the most important one is the high cost. This is the case of the KAT-mobility method consisting of repeating the clusters configuration and calculating the new solution through the set of surviving nodes after detecting failures in any part of the network. It is therefore necessary to adopt a different strategy. Rather than re-optimizing each successive exemplary, we can try to determine a priori solution of the initial problem that can be successively modified in a simple way to solve the following copies. We call this strategy "a priori strategy" [23].

B. Proposed a priori Strategy

A sensor network is modeled by a graph G, and we are operating in a situation where the vertices do not exist in a deterministic way in this graph, but they are present in a probabilistic manner. In other words, a probability of presence will be associated with each vertex. This work is located in a study framework of combinatorial optimization problems when their proceedings are changing in a probabilistic way [10][24]. The aim is to present a priori strategy which ensures the reliability of data transmission in the presence of faults: on any instance of the problem, we avoid the total flaw of the system by changing the graph structure (transformation into a subgraph) according to a modification strategy that will be specified in advance. In fact, we associate the probability p_i to each vertex m_i ∈ V (the probability of remaining operational) taking into account that a center is considered as absent when all its cluster members are faulty. By applying the KAT-mobility protocol, finding a route for the mobile sink is analogous to the traveling salesman problem. A very natural probabilistic extension of this problem was introduced in 1985 for the first time by "Jaillet", when he assumed that the number of cities is a random variable [10]. Concerning the routing path optimization, we propose that only some centers among the K vertices will really require a visit according to their probabilities of occurring. In other words, we specify a strategy µ, called modification strategy, which removes absent centers from the initial a priori tour. Our purpose is to obtain a tour among the initial vertices such that the graph G is transformed into the subgraph G'= G [V'] where V' ⊆ V is the set of present cluster centroids and the new route through its vertices will be in the same order as that established by the a priori tour [25]. This route is illustrated in Figure 2.

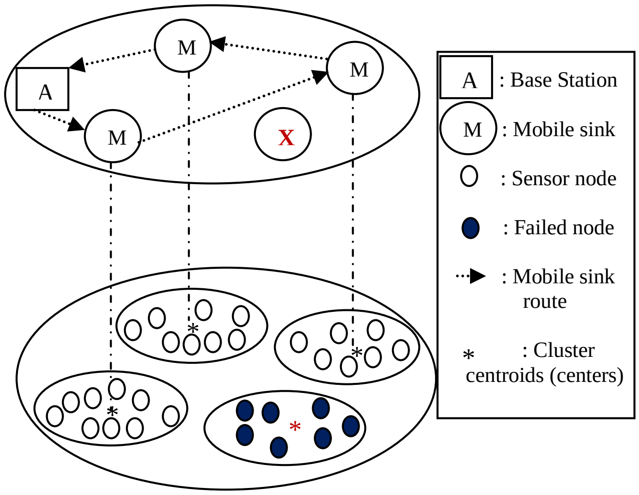


Figure 2. New routing path

This is a problem of finding a priori tour that minimizes the functional of covered distances [18]. Given the probability law ℙ, the set of cluster centroids, the set of all the subsets of V, i.e., each instance V' ⊆ V has a probability of presence ℙ (V'). For a given tour R through the vertices defined on V, the modification method µ consists in deleting or gumming those who are absent from the a priori tour. Let $L_{(R,\mu)}$ be the random variable defined on 2^V, which for all V' of 2^V and with a tour R, associates the length $L_{(R,\mu)}$ (V') through V', induced of the tour R by the modification method. Consequently, the path optimization of the mobile collector to visit once and just once every sensor is equivalent to find the trajectory that minimizes the functional of $L_{(R,\mu)}$ [26][27].

$$\min_{R} \left(\mathbb{E}\,(L_{(R,\mu)}) = \sum_{V'\subseteq V} \mathbb{P}(V')\ L_{(R,\mu)}(V') \right) \quad (5)$$

The use of this a priori strategy allows the wireless sensor network to collect effectively detected information from external environments and deliver it to the required applications with reducing energy consumption, even if some nodes are destroyed. It was proposed to ensure transmission reliability, to increase the network lifetime and also to offer a real time solution, valid in all situations. To ameliorate this work, we propose the improvement of the route optimization in a particular case of this probabilistic problem which is the deterministic case (p_i=1) by implementing the local search method "Tabu" instead of the used algorithm 2-Opt.

V. Simulation Results and Performance Analysis

In this section, simulation results are presented and analyzed. We simulated the KAT-mobility protocol and the proposed approach, based on the optimization of the routing path by using the Tabu algorithm, to evaluate their performance. We deploy 100 sensor nodes which are uniform-randomly distributed inside the simulation area. We consider the metric trajectory cost (expressed in meter) for evaluating the tour lengths. To perform our simulations, we used the software Java.

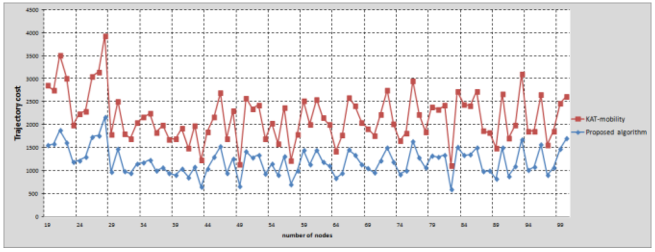


Figure 3. Trajectory cost

As can be seen from the simulation results of Figure 3, which presents the trajectory cost depending on the number of sensor nodes, randomly distributed, that varies from 19 to 100, our improved algorithm achieves better performance than the KAT-mobility algorithm because it guarantees the minimization of the tour length for any employed distribution.

Since the number of nodes may be disturbed and can vary from day to the other, we have proposed our a priori strategy shown at the previous section and to confirm the performance of this probabilistic model, we realized the comparison of the two algorithms.

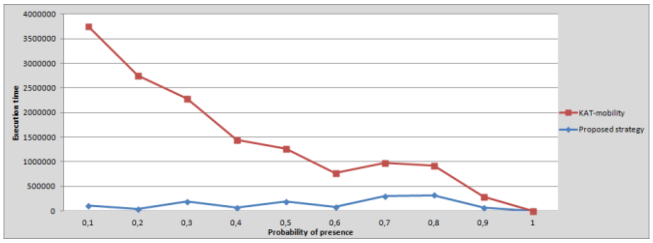


Figure 4. Execution time

It is shown in the simulation results of Figure 4, which presents the time execution (expressed in nanoseconds) depending on the probability of presence of the vertices, that our proposed strategy ensures the obtaining of solutions in real time once the problem is disturbed by the breakdowns of some sensors (the probability of remaining operational is less than 1). This proposed model is much more realistic and provides less execution time than the conventional strategy. It is clear from our simulations that the improvement provided by this new strategy is effective and that this new algorithm is valid in all situations.

VI. Conclusion and Future Works

The data routing in wireless sensor networks is considered as a complex problem because it is essential to ensure reliable delivery of information while consuming less energy. We have presented then in this paper the fault tolerant KAT- mobility protocol, as well as a proposed theoretical probabilistic modeling by considering the fault tolerance as a probabilistic combinatorial optimization problem. Our proposed strategy aims to provide not only better energy efficiency within the wireless network, but also better reliability compared with the conventional approach based on K-means clustering and the approximate solution for Traveling Salesman Problem, especially in the presence of faults. It is shown in the simulation results that this strategy was proposed to provide a better and practical solution by improving the route optimization of the collector. To accomplish this work, we aim to use new methods that navigate the mobile sink to go through the cluster centers according to the optimized route by implementing the "branch & bound" technique instead of the used algorithm 2-Opt, and it will be so interesting when the risk of sensor failures becomes important. Our goal is to propose strategies that guarantee the obtaining of solutions in real time once the problem is disturbed by the failures of some nodes.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” Computer Networks. vol. 38, issue 4, 2002, pp. 393–422, doi: 10.1016/S1389-1286(01)00302-4.

[2] M. Abdallah, J. Bahi, and A. Mostefaoui, “Sensor Networks: localization, coverage and data fusion,” Franche-Comté, 14 -11-2008.

[3] C. Karlof and D. Wagner, “Secure routing in wireless sensor networks: attacks and countermeasures,” Ad Hoc Networks, vol. 1, 2003, pp. 293–315, doi: 10.1016/S1570-8705(03)00008-8.

[4] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, “Fault-resilient sensing in wireless sensor networks,” Computer communications archive, vol. 30, issues 11–12, 2007, pp. 2375–2384, doi: 10.1016/j.comcom.2007.04.023

[5] S. Mishra, L. Jena, and A. Pradhan, “Fault tolerance in wireless sensor networks,” International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, issue 10, October 2012, pp. 146-153.

[6] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, “Coverage problems in wireless ad-hoc sensor networks,” IEEE Infocom2001, Ankorange, Alaska, April 2001, pp. 1380-1387, doi: 10.1109/INFCOM.2001.916633.

[7] Y. Challal, A. Ouadjaout, N. Lasla, M. Bagaa, and A. Hadjidj, “Secure and efficient disjoint multipath construction for fault tolerant routing in wireless sensor networks,” Journal of network and computer applications, vol. 34, issue 4, July 2011, pp. 1380-1397, doi: 10.1016/j.jnca.2011.03.022.

[8] M. Bellalouna and A. Ghabri, “A priori methods for fault tolerance in wireless sensor networks,” World Congress on Computer and Information Technologies (WCCIT), IEEE, Sousse, June 2013, pp. 1-6, doi: 10.1109/WCCIT.2013.6618654.

[9] W. Tekaya, V. T. Paschos, and C. Murat, “Minimim probabilistic spannig tree problem,” Université Paris Dauphine, 2008.

[10] M. Bellalouna, “Probabilistic combinatorial optimization problems,” Ph.D thesis, Ecole nationale des ponts et chaussées, Paris, 1993.

[11] J. P. Jafrin and P. A. Christy Angelin, “A comparative study of data gathering algorithms for a mobile sink in wireless sensor network,” International journal of advanced research in computer engineering and technology, vol. 1, issue 9, November 2012, pp. 250-255.

[12] H. Nakayama, N. Ansari, A. Jamalipour, Y. Nemoto, and N. Kato, “On data gathering and security in wireless sensor networks,” University of Sydney, Sarnoff symposium, IEEE, 2006, pp. 1-7, doi: 10.1109/SARNOF.2007.4567337.

[13] J. Kogan, “Introduction to Clustering Large and High-Dimensional Data,” Cambridge University Press, Cambridge, 2007.

[14] Z. Guellil and L. Zaoui, “Proposition of a solution to the initialization problem K-means Case,” CIIA, CEUR Workshop Proceedings, CEURWS. Org, vol. 547, 2009, pp. 1-9.

[15] P. Mathur, S. Saxena, and M. Bhardwaj, “Node clustering using K Means Clustering in Wireless Sensor Networking,” 2nd National Conference in Intelligent Computing Communication, Dept. of IT, GCET, Greater Noida, INDIA, 2013, pp. 1-6.

[16] T. Fukabori, H. Nakayama, H. Nishiyama, N. Ansari, and N. Kato, “An efficient data aggregation scheme using degree of dependence on clusters in WSNs,” Communications (ICC), IEEE International Conference, May 2010, pp. 23-27, doi: 10.1109/ICC.2010.5502285.

[17] M. A. Chikh, M. Feham, H. Guyennet, A. Benyettou, and M. Lehsaini, “Diffusion and coverage based on clustering in sensor networks,” Université de Franche-Comté, 2009.

[18] P. Jaillet, “The Probabilistic Traveling Salesman Problems,” Technical report 185, Operations research center, MIT, Cambridge, Mass, 1985.

[19] D. S. Johnson, L. A. Mcgeoch, and E. E. Rothberg, “Asymptotic experimental analysis for the held-karp traveling salesman bound,” in proceedings of the 7th annual ACM-SIAM symposium on discrete algorithms, Atlanta, Georgia, January 1996, pp. 341-350, doi: 10.1.1.47.7327.

[20] I. Or, “Traveling salesman-type combinatorial problems and their relation to the logistics of regional blood banking,” Ph. D. thesis, Northwestern University, Evanston.

[21] D. Johnson and L. McGeoch, “The Traveling Salesman Problem: A Case Study in Local Optimization,” John Wiley & Sons, 1997.

[22] N. Boria, C. Murat, and V. Th. Paschos, “An emergency management model for a wireless sensor network problem,” CAHIER DU LAMSADE 325, Juillet 2012, pp. 1-23.

[23] D. Bertsimas, P. Jaillet, and A.Odoni, “A priori Optimization,” Operations Research, vol. 38, 1990, pp. 1019-1033, doi: 10.1287/opre.38.6.1019.

[24] M. Bellalouna, C. Murat, and V. Paschos, “Probabilistic combinatorial optimization problems on graphs: A new domain in operational research,” European Journal of Operational Research, vol. 87, issue. 3, 1995, pp. 693-706, doi: 10.1016/0377-2217(95)00240-5.

[25] C. Murat and V. Paschos, “probabilistic combinatorial optimization on graphs,” Wiley- ISTE, London, 2006.

[26] P. Jaillet, “A priori solution of a Travelling Salesman Problem in which a random subset of the customers are visited,” Operations research, vol. 36, issue 6, 1988, pp. 929-936.

[27] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, “The Traveling Salesman Problem,” John wiley & sons, 2011, doi: 10.1002/9781118033142.ch7.

Wireless Ad Hoc Networks Resilience Through Cooperative Communication

Ulisses Rodrigues Afonseca
Instituto Federal de Goiás - IFG
Campus Luziânia
Luziânia-GO, Brazil
urafonseca@ifg.edu.br

Thiago Fernandes Neves, Jacir Luiz Bordim
Department of Computer Science
University of Brasília - UnB
Brasília-DF, Brazil
tfn.thiago@cic.unb.br, bordim@unb.br

***Abstract*—Several studies predict the use of wireless ad hoc networks in support of critical missions like search and rescue and prevention of natural disasters. In order to improve network connectivity, Cooperative Communication (CC) has been explored as an alternative to connect isolated network components in wireless ad hoc settings. However, existing cooperative communication solutions rely on global topological information which may not be feasible in more realistic scenarios. This paper presents a self-organized and distributed solution to improve network connectivity by exploring the availability of cooperative communication links. Simulation results show that the proposed scheme has a low computation cost and provides a link recovery rate comparable to those obtained by centralized solutions.**

***Keywords*–Ad hoc networks; articulation; bridge; cooperative communication; critical edge; critical node; network resilience.**

I. Introduction

In an ad hoc network, nodes cooperate in relaying packets to each other to enable communication. In such scenarios, network connectivity is crucial once faraway nodes may not be able to communicate in case of network partitioning. Link and node failure are events that may occur during the course of operation. As urgent and critical tasks, such as search and rescue and the prevention of natural disasters depend on network connectivity, ways to prevent network partitioning and node isolation are of interest. Let $G = (V, E)$ be a undirected connected graph, where V is the set of nodes and E is the set of edges. A node $v \in V$ is an *articulation* or *cut-vertex* if its removal makes the graph disconnected. Similarly, an edge $e \in E$ is a *bridge* if its removal makes the graph disconnected. Note that bridge links are connected by two articulations nodes. In this work, articulation nodes sharing a bridge are referred to “bridge nodes”.

Owing to their importance in preserving network connectivity, ways to identify articulation nodes and bridges has been investigated in the literature. Goyal and Caffery Jr [1] proposed a centralized mechanism to identify articulations in wireless networks. Later, Jorgic et al. [2] presented a distributed solution, where each node performs a k-hop depth-first search to locate and identify articulation nodes and bridges using localized information. The proposed solution, however, has the drawback of false positive detections. Chaudhuri [3] and Turau el at. [4] proposed algorithms based on distributed depth-first search to determine the articulations of a graph. These solutions are optimal in time and number of messages and work with the knowledge of directly connected neighbours only.

As viable solutions to locate bridges and articulation nodes have been developed, the research community focused on alternatives to extend the availability of such nodes and links as well as in ways to reestablish network connectivity in case of failure [2][3][4]. Afonseca et al. [5] proposed ways to reduce energy consumption of articulation nodes by using packet aggregation techniques. The solution is based on the fact that energy consumption of articulation nodes is usually higher than other nodes, leading to a premature node failure. Khelifa et al. [6] propose the usage of dormant nodes that could be activated in case of link or node failure. When necessary, the dormant nodes would be activated to prevent network partitioning. In the same line, Goyal and Caffery Jr [1] proposed the usage of limited, coordinated, mobility so as to recover from link and node failures. In case of network disruption, nodes would move in a coordinated way as to reestablish network connectivity. As the location of articulation nodes are assumed to known a priory, node activation and dispatch can be employed to improve network connectivity in such areas. Yu et al. [7] employed *cooperative communication* techniques to connect disjoint network components using cooperative links. The proposed technique allows transmitting nodes to transpose the limit of maximum transmission range by allowing multiple nodes to relay the same information, thus improving network connectivity [7][8]. Neves et al. [9] exploited cooperative communications to establish power efficient links and routes to a sink node.

Despite of its benefits, all the above solutions rely on global topological information which may not be feasible to obtain and maintain due to its operational costs. Also, it seems unlikely that a network would have enough dormant nodes with the ability to be activated and dispatched to the necessary location whenever needed. Hence, cooperative communication seems to be a suitable approach to improve network connectivity. However, the aforementioned works that explore this path aimed to locate the least power link cost that can connect disjoint components before operation. This work takes a different approach by focusing on localized mechanisms to prevent network partitioning due to node and link failure. More precisely, the proposed scheme works by identifying critical elements, such as bridges and articulation nodes, and uses cooperative communication to create cooperative links to avoid network partitioning whenever possible. The proposal scheme employs cooperative communication based on localized topology information and works in a distributed, self-organizing, manner. Simulation results show that the proposed scheme has a low computation cost while link recovery rate is comparable to those obtained by centralized solutions.

The rest of the paper is organized as follows: Section II presents a review on cooperative communication, defines the communication model and formalizes the problem addressed in this work. Section III presents a distributed solution to recover network connectivity. In Section IV, the simulation process and the data collected are presented, and finally, Section V concludes this paper.

II. Communication Model and Problem

Cooperative Communication (CC) aims at enabling the cooperation of nodes for transmitting their messages to the destination [7][8]. Rather than operating independently, competing with each other for channel resources, nodes form a virtual Multiple Input Multiple Output (MIMO) system and simultaneously transmit the same information. In MIMO, nodes use a set of antennas to transmit and receive data to combat signal fading. Cooperative communication is similar, but uses multiple nodes in a two-step process [7][10]. In the first step, a node, called source, sends information for a subset of the nodes directly connected, called “helper nodes". In the second step, the source and helpers send the same data simultaneously. Thus, cooperative communication enjoys the same benefits of a conventional MIMO system. Further details on the characteristics of the cooperative communication can be found in Hong et al. [8]. The subsequente sections describe the cooperative communication model considered, present a brief overview of the closely related works and formally defines the problem addressed in this work.

A. Cooperative Communication Model

The communication between two nodes, in the traditional model, can be simplified in terms of the transmission power, the distance between nodes and the rate of signal fading. Thus, consider a network modelled as a planar, undirected graph $G(V, E)$, where $V = \{v_1, v_2, ..., v_n\}$ is a set of wireless nodes and E is the set of communication links. Each node v_i can adjust its transmit power p_i with values in the range $[0, P_{MAX}]$. When $p_i = 0$, the transceiver is turned off and when $p_i = P_{MAX}$, the transceiver operates at full power. In traditional models of communication, the source v_i can communicate directly with the destination node v_j only when the transmission power of v_i complies with (1):

$$P_i(d_{i,j})^{-\alpha} \geq \tau \qquad (0 \leq P_k \leq P_{MAX}), \tag{1}$$

where α is the exponent of signal fading, usually around 2 and 4, which is the rate of loss of the signal power with increasing distance, $d_{i,j}$ is the Euclidean distance between v_i and v_j, and τ is the receiver sensitivity to correctly receive a packet, i.e., the threshold of the received power so that node v_j can correctly decode the signal and obtain the original message.

In cooperative communication, the transmission power required by the source in conjunction with helper nodes can be determined similarly to the direct communication. Full communication between nodes v_i and v_j can be obtained with CC, if v_i transmits its signal with an auxiliary node set $H_{i,j}$ and the sum of transmission powers satisfies (2).

$$\sum_{v_k \in v_i \cup H_{i,j}} P_k(d_{k,j})^{-\alpha} \geq \tau \qquad (0 \leq P_i \leq P_{MAX}). \tag{2}$$

In the cooperative communication model, new concepts are introduced whereas the new edges in the graph can not be defined using classical concepts. Then, some important definitions on the model of cooperative communications, similar to those presented by Zhu et al. [10], are shown:

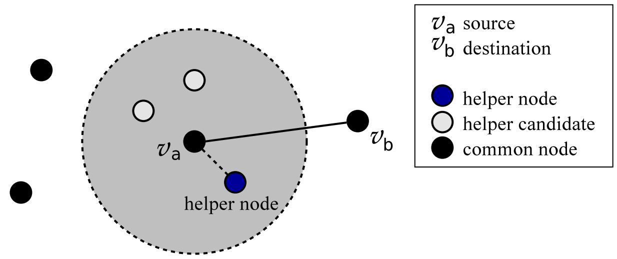


Figure 1. Example of the cooperative communication.

Definition 2.1: (Direct link): A direct link $\overline{v_i v_j}$ is an edge in E representing that the node v_i can transmit data to node v_j directly, that is, p_i is such that the node v_i can reach v_j when $p_i \leq P_{MAX}$. A solid horizontal line on the nodes represents a direct link.

Definition 2.2: (Helper node set): $H_{i,j}$ is the set of helper nodes of v_i in a cooperative communication with v_j. It is assumed that all required helper nodes are direct neighbours of v_i, that is, $H_{i,j} \subseteq N(v_i)$, where $N(v_i)$ is the set of all direct neighbours of v_i. In other words, all elements in $N(v_i)$ are candidates for helper nodes.

Definition 2.3: (CC-link): A CC-link $\widetilde{v_i v_j}$ is an edge of E representing that a node v_i can transmit data to v_j cooperatively using a set of auxiliary nodes $H_{i,j}$. A horizontal wavy line is used to denote a CC-link.

Definition 2.4: (Helper link): A helper link is an edge between v_i and one of his helpers in $H_{i,j}$.

Definition 2.5: (Network topology): The union of all direct links and CC-links, $\overline{E}$ and $\widetilde{E}$, respectively. Similarly, the graph of direct communication and CC communication are denoted by $\overline{G} = (V, \overline{E})$ and $\widetilde{G} = (V, \widetilde{E})$, respectively. Note that $E = \overline{E} \bigcup \widetilde{E}$. Also, if $v_i v_j \in E$, then: $v_i v_j = \overline{v_i v_j}$ if $v_i v_j$ is a direct link; and $v_i v_j = \widetilde{v_i v_j}$ if $v_i v_j$ is a CC-link.

Definition 2.6: (Weight of direct link): The weight of a direct link $\overline{v_i v_j}$ is defined as: $w(\overline{v_i v_j}) = \tau d_{i,j}^{\alpha}$.

Definition 2.7: (Weight of a CC-link): The weight of a CC-link $\widetilde{v_i v_j}$ is defined as:

$$w(\widetilde{v_i v_j}) = w_d(H_{i,j}) + (|H_{i,j}| + 1) w_{CC}(H_{i,j}),$$

where:

- $|H_{i,j}|$: is the number of elements in $H_{i,j}$;
- $w_d(H_{i,j}) = \left(\frac{\tau}{\max_{v_k \in H_{i,j}} (d_{i,k})^{-\alpha}}\right)$: is the maximum power consumption of the node v_i to communicate with the farthest node in $H_{i,j}$;
- $w_{CC}(H_{i,j}) = \left(\frac{\tau}{\sum_{v_k \in v_i \bigcup H_{i,j}} (d_{k,j})^{-\alpha}}\right)$: is the minimum power consumption of the node v_i to communicate directly to v_j, together with their helper nodes in $H_{i,j}$

In a cooperative communication from v_i to v_j, node v_i should initially send its data to helper nodes in $H_{i,j}$ and then, node v_i and its helpers must simultaneously send the same data to v_j. Thus, the weight of a CC-link is the sum of communication cost of these two steps. The cost for the first stage of communication is equivalent to $w_d(H_{i,j})$, while the cost of individual nodes to transmit data using CC is $w_{CC}(H_{i,j})$. Figure 1 shows a cooperative communication example. The radius of maximum transmission, represented in a grey circle, shows that node v_a has three neighbours. To communicate with the destination node v_b, out of its reach, the source node v_a uses a helper candidate sharing a direct link to v_a.

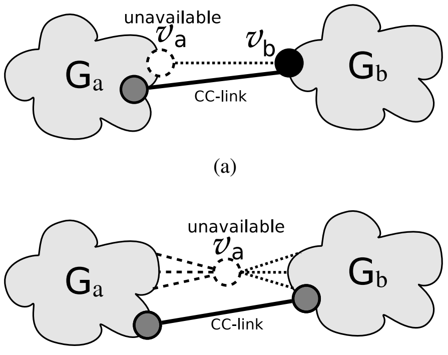
B. Related Works

In wireless ad hoc networks, CC have been used as a topology control mechanism with the aim to improve network connectivity and while reducing power consumption [7][10][11]. For networks initially without full connectivity, CC was used to transpose the maximum transmission range as a mean to improve network connectivity [7][9]. Yu et al. [7] use CC as a topology control mechanism, whose purpose is to connect disjoint components through cooperative links. The proposed solution, called CoopBridges, increases the network connectivity while reducing the transmission power at each node. The authors proposed a heuristic to select power efficient helpers nodes to reduce power consumption of the nodes sharing a CC-link. Starting from an undirected, disconnected graph, CoopBridges uses the proposed heuristic to create cooperative edges to connect components in the network. In the resulting topology, the minimum spanning tree algorithm is employed within each component and between network components to prune costly links. The task of selecting power efficient helper nodes works in $O(|V|^2)$ time. Neves et al. [9] developed a similar mechanism that interconnects the components of an ad hoc network, initially with no direct connectivity with a sink node. The solution consists of four steps that uses a modified version of the heuristic proposed by Yu et al. [7]. The solution uses low cost cooperative edges to interconnect the network such that paths created should lead to the sink node. The task of selecting power efficient helper nodes takes $O(|V|^2)$ time [9].

This work presents a localized and proactive strategy to maintain network connectivity in the event of network partitioning. Unlike the previous works, that are based on coordinated mobility or dormant nodes, the proposed scheme uses cooperative communication. Localized information is used to reduce the computational cost to select helper nodes during the process of establishing cooperative links. To the best of our knowledge, this is the first work to employ cooperative communication in a proactive way to prevent network partitioning.

C. Problem Formulation

This paper address the problem of recovering network connectivity in the event of node and link failure on ad hoc networks with cooperative communication capabilities similar to those in [7][8][10][12][13]. In particular, this work focuses on monitoring articulation points and bridges nodes and, in case of unavailability, cooperative communication is employed as to reestablish network connectivity. Consider an ad hoc wireless network represented by a planar, undirected graph $G(V, E)$ such that there are a number of articulation points and bridges. By definition of articulation point and bridge, on the event of unavailability of one of these elements, the graph G becomes disconnected. Let G_a and G_b be the components of G that have been created due to the unavailability of an articulation point or a bridge node. Furthermore, let node $v_a \in G_a$ and $v_b \in G_b$ and let $H(v_a)$ and $H(v_b)$ be the set of helper nodes available to v_a and v_b, respectively. Thus, using cooperative communication, the proposed scheme aims to reconnect the graph G. In this context, the problem addressed in this work is three fold: (i) locate articulations and bridges in the network; (ii) monitor their status and; (iii) in case of unavailability, coordinate the activities of the neighbouring nodes to recovery connectivity with the aid of cooperative communication.



(a)

(b)

Figure 2. Representation of the problem of connectivity recovering using CC after a (a) bridge edge failure and (b) an articulation node failure.

III. Proposed Solution

This section presents a distributed algorithm that allows the network to reestablish connectivity in case of articulation node failure. The main idea is to proactively identify suitable CC-links and to ensure that these CC-links are created in case of network connectivity disruption. Figure 2 shows an example of connectivity recovery in cases of bridge and articulation failure. The dotted edges represent topology changes that effect the communication links. When a bridge node fails, according to Figure 2a, collaborative communication is used to recover connectivity by establishing a cooperative link among the remaining bridge node and nodes in the vicinity of the failed bridge node. When there is only one articulation node, identified by v_a in Figure 2b, that connects two components, collaborative communication is employed as an alternative to reconnect the graph.

To achieve the above, the proposed scheme uses two-hop information to allow articulation nodes to periodically update their neighbours so that they can create CC-links to maintain network connectivity in the case of articulation node failure. To perform power efficient selection of helper nodes, the Greed Helper Set Selection ($GHSS$) heuristic, proposed by Yu et al. [7], is employed as a routine in the main algorithm, similarly to [10]. A call to the heuristic $GHSS$ takes as input parameters the pair (v_s, v_d), where v_s is the source node and v_d is the destination node, the output of this call is the cost of the CC-link $\widetilde{v_s v_d}$ or ∞ if it is not possible to create the CC-link. Note that, despite the related works that propose the increase of network connectivity using CC, the goal of the proposed solution focus on connectivity recovery. Another important aspect of the proposed solution is the processing type and amount of information used to reconnect the network. The proposed solution is distributed and uses localized information, while other proposals in the literature are centralized and require global topological information. The subsequent section details the proposed scheme.

A. Reconnecting Components (RC)

The proposed solution, called *ReconnectComponents* (RC), is detailed in Algorithm 1 (Figure 3). The algorithm considers that each node knows (i) whether it is part of a bridge or not,

Algorithm 1 RC($articulation$, $bridge$, S)

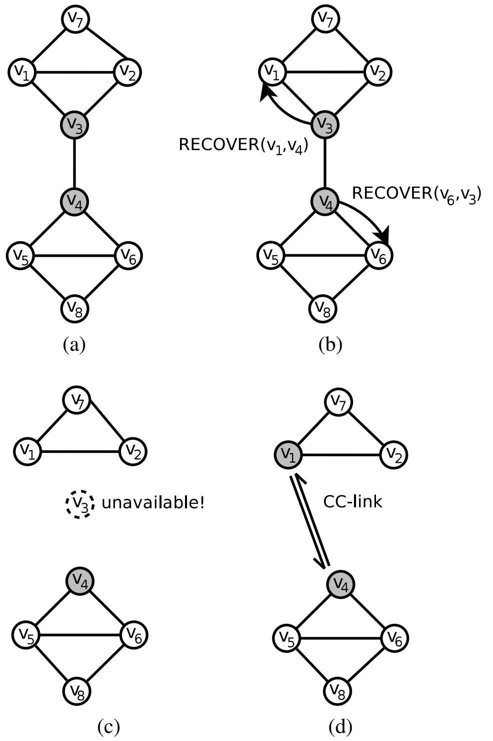
```
# Articulation node forming a bridge notifies a
# neighbouring nodes to replace it
 1: if ($articulation = TRUE$ and $bridge = TRUE$) then
 2:     Let $v_a$ and $v_b$ be the nodes sharing a bridge. $G_a$ and $G_b$ are the
        network component connected to $v_a$ and $v_b$, respectively;
 3:     Let $N_{v_a}(G_a)$ be the set of neighbours of $v_a$ in $G_a$;
 4:     for each $S$ seconds do
 5:         $v_a$ computes $GHSS(v_i, v_b)$ and $GHSS(v_i, v_b)$, $\forall v_i \in$
            $N_{v_a}(G_a)$, and finds $v_k$, $v_k \in N_{v_a}(G_a)$, such that the combined
            cost of CC-links $\widetilde{v_k v_b}$ and $\widetilde{v_b v_k}$ are minimum;
 6:         $v_a$ send $RECOVER(v_k, v_b)$ to nodes $v_k$ and $v_b$;
 7:     end for
 8: end if

# Articulation node computes the CC-link cost to
# connect its neighbouring nodes using local information
 9: if ($articulation = TRUE$ and $bridge = FALSE$) then
10:     Let $v_a$ be an articulation node running the algorithm;
11:     Let $G_i$ and $G_j$ be two network components that are connected to $v_a$
        such that $G - v_a$= $G_i \bigcup G_j$;
12:     Let $N_{v_a}(G_a)$ be the set of direct neighbours of $v_a$ in component
        $G_a$;
13:     for each $S$ seconds do
14:         $v_a$ computes $GHSS(v_i, v_j)$ and $GHSS(v_j, v_i)$, $\forall v_i \in$
            $N_{v_a}(G_a)$ and $\forall v_j \in N_{v_a}(G_b)$, and find $v'_i$ and $v'_j$, $v'_i \in$
            $N_{v_a}(G_a)$ and $v'_j \in N_{v_a}(G_b)$, such that the combined cost to
            create the CC-links $\widetilde{v'_i v'_j}$ and $\widetilde{v'_j v'_i}$ are minimum;
15:         $v_a$ send $RECOVER(v'_i, v'_j)$ to nodes $v'_i$ and $v'_j$;
16:     end for
17: end if

# Actions taken by nodes receiving a RECOVER msg
18: Let $\{v_i, v_j\}$ be the set of nodes receiving a $RECOVER(v_i, v_j)$
    message;
19: Let $v_a$ be the articulation node the sent the $RECOVER$ message;
20: while true do
21:     if $v_a$ is unavailable then
22:         Create the CC-links $\widetilde{v_i v_j}$ and $\widetilde{v_j v_i}$;
23:     end if
24: end while
```

Figure 3. Algorithm Reconnect Components

and (ii) whether it is an articulation or not. This knowledge can be obtained by running algorithms such as those proposed in [2][3]. Besides these requirements, the algorithm takes as input a parameter S that indicates the time interval in which the cooperative links are computed and updated to accommodate eventual topological changes.

Bridge nodes calculates the bidirectional cooperative link having the least cost between the adjacent articulation and its direct neighbours (lines 1-5). A message is sent to the elected nodes that compose the cooperative link (line 6). Nodes receiving the $RECOVER(v_i, v_j)$ message should monitor the source of the message (lines 18-24). Should the articulation node become unavailable, the cooperative link is then used to maintain connectivity (line 22).

When a node is an articulation and does not compose a bridge, it calculates the cooperative link with least power cost between each pair of neighbouring nodes in the components interconnected by it (line 9-14). After that, a message is sent to the selected nodes (line 15). Note that, up to this point, the articulation only elects nodes to replace it and the links supported by the articulation node. It should be noted that after an articulation node failure, that is not associated to a bridge, a new bridge is created and therefore two new bridge nodes appear in the graph. In this case, the connectivity can be maintained continuously using the same strategy.



(a) (b)

(c) (d)

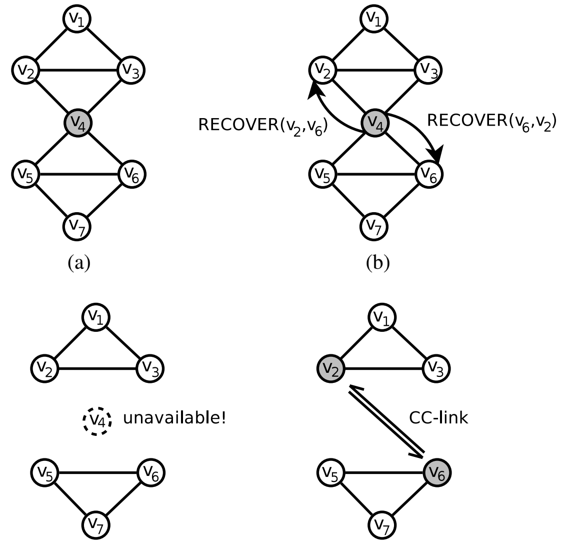
Figure 4. Example of connectivity recovery after a bridge node failure.

B. A Working Example

Figure 4 and 5, respectively, show the sequence of events for recovering the connectivity when a bridge and an articulation become unavailable. Figure 4a presents the initial topology in which nodes v_3 and v_4 are bridge nodes. In Figure 4b, these nodes notify neighbours that should take over its function in case of failure. In Figure 4c, articulation v_3 fails and in Figure 4d a cooperative link is created between nodes v_1 and v_4. In Figure 5b, the articulation node v_4 notify nodes v_2 and v_6 that they have been elected to create a cooperative link. In Figure 5c, when the articulation v_4 fails, a cooperative link is created and the new topology is presented in Figure 5d. In the resulting topology, there are two new articulation nodes and a bridge.

C. Computational Cost

The $GHSS$ routine is used to perform the required helper selection in the RC algorithm (lines 5 and 14) and is also used, as in previous works, to measure the computational cost of the proposed algorithm. For this purpose, let $\Delta(G)$ denote the maximum degree of a node in G. Also, let v_a and v_b denote two nodes connected by a bridge. According to previous definitions, node v_a and v_b have $N_{v_a}(G_a)$ and $N_{v_b}(G_b)$ neighbours, respectively. Note that $N_{v_a}(G_a) \subset N(v_a)$ and $N_{v_b}(G_b) \subset N(v_b)$. Then, the task of computing the best set of helper nodes to create CC-links (line 5) makes at most $2[N(v_a) + N(v_b)]$ calls for the $GHSS$ routine. In the case where a node v_a is an articulation node such that $G - v_a$= $G_i \bigcup G_j$, node v_a computes the best set of helper nodes among its one and two-hop neighbouring nodes. Hence, in the worst case, the direct neighbours of v_a, say $v_i \in N(v_a)$ has a degree of at most $\Delta(G)$. As the heuristic to compute the best set of helper nodes needs to verify all alternatives among the



(a) (b) (c) (d)

Figure 5. Example of connectivity recovery after a failure in an articulation.

nodes in $N(v_i)$ that connect to $v_j \in G_j$ and vice-versa, node v_a makes, in the worst case, $O(\Delta(G)^2)$ calls to the $GHSS$ routine. Considering that $2[N(v_a) + N(v_b)] \leq O(\Delta(G)^2)$, the RC algorithm uses, in the worst case, $O((\Delta(G))^2)$ calls to the $GHSS$ to select the best helper set to establish a CC-link.

IV. Evaluation and results

The proposed solution has been evaluated by simulation. The validation process consists in: (i) generate random topologies; (ii) identify articulation nodes and bridge nodes; (iii) employ the RC algorithm to compute the best CC-links and check whether these links are able to reconnect the network in case of articulation and bridge nodes failure. To assess the goodness of the proposed solution, the resulting CC-link cost is compared with those produced by the centralized scheme presented by Yu et al. [7].

The evaluation scenarios are based on the following parameters (similarly to those in [9][10][13]): a set of nodes $n = 20, 30, ..., 60$ are randomly positioned in a 300×300m area. Equation (1) can be easily adapted to a more suitable path loss models. Hence, in what follows, Free Space Path Loss Model [14] is considered, where the maximum transmitting power (P_{MAX}) is set to 6dBm and the receiver threshold (τ) is set to -71dBm, allowing to a maximum transmission range (R_{MAX}) of $\approx$ 70m on the 2.4GHz frequency band [14]. To compare the performance of the proposed solution, the following metrics were considered:

- **M1:** **Computational cost:** Aims to evaluate the overhead (in terms of calls to $GHSS$) to select the helper set with least power cost to create CC-link using distributed and a centralized solutions;
- **M2:** **Power cost:** The amount of transmission power needed to establish the CC-links;
- **M3:** **Percentage of recovered connectivity**: identifies the percentage of graphs that had connectivity recovered;

To compute **M1**, first a random topology is generated using the defined parameters. Then, the articulation nodes and bridges are identified in the graph and the best helper sets are computed. For a defined node density, the simulation results are drawn from an average of a three hundred random topologies. Note that, as the purpose of **M1** is to compare the computational cost, the parameter S has no effect in this case. Table I presents the simulation results for metric M1. The column "density" corresponds to the number of nodes in the graph. Columns "Articulations" and "Bridges" correspond to the number of calls to the $GHSS$ heuristic. The column "Global" shows the same results when the best cooperative bi-direction link among all the nodes is computed. Note that the degree of each node is random. Thus, as the node density gets higher, nodes tend to have a larger node degree and, consequently, increasing the computational cost. However, even by increasing the degree of the graph, the proposed solution still has a scalable computational cost. As can be seen in the Table, the proposed algorithm obtained, for the evaluated cases, a reduction of up to 67 times the number of calls to the GHSS routine.

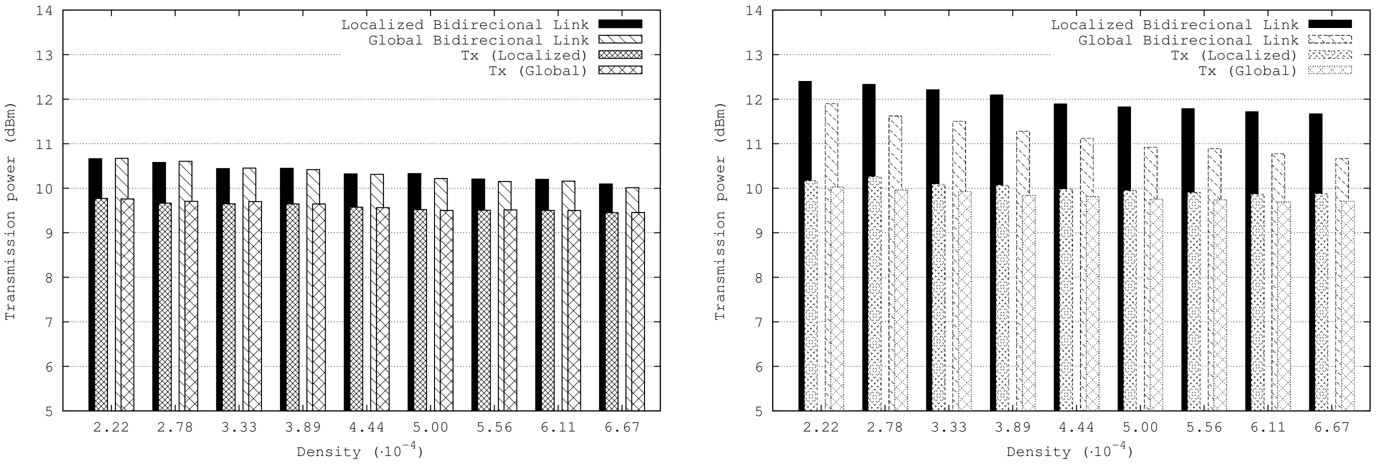
The simulation results for metric **M2** are shown in Figure 6. In the figure, the x-axis represents the graph density (number of nodes per area) and the y-axis represents the power consumption for the computed CC-links. As can be seen in Figure 6a, the proposed algorithm has an power consumption slightly higher than then that provided by the global bi-directional link when an articulation connecting two components fails. This trend is also verified in Figure 6b that shows the power consumption to reestablish connectivity in case of an bridge node failure. The average power consumption reduces as the network density increases. This occurs as the articulation and bridge nodes have potentially more nodes that may act as helper nodes. With closer helper nodes, the cost $w_d(H_{i,j})$ decreases and the weight of the cooperative link $w(\widetilde{v_i v_j})$ tends to reduce as well. Despite the slightly higher power consumption of the proposed solution, it is important to state that the RC relies solely on local information.

Suppose that the nodes could increase the transmission power beyond the P_{MAX} up to the limit necessary to reestablish network connectivity without the aid of a CC-link. Although this is an unlikely situation, it provides a lower bound on the minimum amount of power necessary to reestablish communication. Also in Figure 6b, the bar `Tx(Localized)` corresponds to the minimum transmission power needed to reestablish connectivity without resorting to CC-links in the resulting topology using localized information. Similarly, the `Tx(Global)` bar corresponds to the amount of power necessary to reestablish connectivity using global information. When a direct link are considered, the nodes selected in both cases present comparable results in terms of the required transmitting power to reestablish network connectivity.

For the metric **M3**, percentage of graphs that had con-

TABLE I. NUMBER OF CALLS TO THE GHSS ROUTINE (M1).

Density ($\times 10^{-4}$)	Articulation	Bridge	Global
2.22	13.22	14.78	174.04
2.78	14.52	16.38	276.98
3.33	15.62	17.18	408.20
3.89	16.74	18.86	562.78
4.44	17.78	21.98	742.00
5.00	18.90	21.36	950.72
5.56	19.26	24.96	1177.96
5.11	20.34	22.48	1431.12
6.67	22.12	25.24	1710.72



(a)

(b)

Figure 6. Average power cost to recover connectivity in case of: (*a*) an articulation node failure; and (*b*) a bridge node failure.

nectivity recovered, it was observed that the proposed solution presents similar results to the global alternative. This results are not shown due to space limitation. Nevertheless, the observed results shows that, for node density up to 2.5, a success rate of 98% by employing global information while RC attains 96.5%, only 1.5 points below the extensive evaluations. For node density with values between 2.5 and 5.0, both algorithms were able to recover network connectivity in approximately 99% of the cases. On graphs with node density above 5.0, both algorithms have been able to reconnect the graphs in all evaluated cases.

V. Conclusion

Maintain connectivity in wireless ad hoc networks is a goal that has been addressed in many ways, most of them focusing on identifying critical nodes and implement mechanisms the preserve these nodes using efficient routing, packet aggregation, among other techniques [2][3][4]. This work explored cooperative communication to reconnect the network using distributed processing and localized knowledge. Bidirectional links are created between network components when articulation and bridge nodes fail. The main contribution of this work is to present an algorithm that reduces the computational cost when using localized information that offers resilience when monitoring critical elements, creating cooperative links when conventional links become unavailable. Simulation results demonstrate that the proposed solution provides similar results of more costly solutions that rely on global topological information. For the scenarios evaluated, the computational cost of the proposed scheme was 67 times lower than centralized solution, while producing comparable results.

Acknowledgment

This work was partially supported by CNPq, CAPES and UnB/DPP.

References

[1] D. Goyal and J. Caffery Jr, “Partitioning avoidance in mobile ad hoc networks using network survivability concepts,” in Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on. IEEE, 2002, pp. 553–558.

[2] M. Jorgić, I. Stojmenović, M. Hauspie, and D. Simplot-Ryl, “Localized algorithms for detection of critical nodes and links for connectivity in ad hoc networks,” Proceedings of the 3rd IFIP Mediterranean Ad Hoc Networking Workshop, 2004, pp. 360–371.

[3] P. Chaudhuri, “An optimal distributed algorithm for finding articulation points in a network,” Computer Communications, vol. 21, no. 18, 1998, pp. 1707–1715.

[4] V. Turau, “Computing bridges, articulations, and 2-connected components in wireless sensor networks,” in Algorithmic Aspects of Wireless Sensor Networks. Springer, 2006, pp. 164–175.

[5] U. R. Afonseca, P. H. Azevêdo Filho, J. L. Bordim, and P. S. Barreto, “Reducing energy consumption of articulation points in wireless sensor networks,” in II Workshop de Sistemas Distribuídos Autonômicos, 2012, pp. 21–24.

[6] B. Khelifa, H. Haffaf, M. Madjid, and D. Llewellyn-Jones, “Monitoring connectivity in wireless sensor networks,” in Computers and Communications, 2009. ISCC 2009. IEEE Symposium on. IEEE, 2009, pp. 507–512.

[7] J. Yu, H. Roh, and W. Lee, “Topology Control in Cooperative Wireless,” IEEE Journal on Selected Areas in Communications, vol. 30, no. 9, 2012, pp. 1771–1779.

[8] C.-C. J. K. Y.-W. Peter Hong, Wan-Jen Huang, Cooperative Communications and Networking: Technologies and System Design. Springer Science & Business Media, 2010.

[9] T. F. Neves and J. L. Bordim, “Topology control in cooperative ad hoc wireless networks,” Electronic Notes in Theoretical Compututer Science, vol. 302, 2014, pp. 29–51.

[10] 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.

[11] M. Cardei, J. Wu, and S. Yang, “Topology control in ad hoc wireless networks using cooperative communication,” Mobile Computing, IEEE Transactions on, vol. 5, no. 6, 2006, pp. 711–724.

[12] M. Uysal, Cooperative Communications for Improved Wireless Network Transmission: Framework for Virtual Antenna Array Applications. Information Science Reference, 2010.

[13] J. Yu, H. Roh, W. Lee, S. Pack, and D.-Z. Du, “Cooperative bridges: Topology control in cooperative wireless ad hoc networks,” in INFOCOM, 2010 Proceedings IEEE. IEEE, 2010, pp. 1–9.

[14] J. C. Liberti and T. S. Rappaport, Smart Antennas For Wireless Communications: IS-95 And Third Generation CDMA Applications, ser. Prentice Hall Communications Engineering And Emerging Technologies Series. Upper Saddle River, N.J.: Prentice Hall PTR, 1999.

Node Pair Selection Scheme for MIMO Multiuser System Using Repeated Application of Stable Matching Algorithm

Testuki Taniguchi and Yoshio Karasawa
The Department of Communication Engineering and Informatics
Advanced Wireless Communication research Center (AWCC)
The University of Electro-Communications (UEC)
Tokyo, 182-8585 Japan
Email: taniguch@ee.uec.ac.jp, karasawa@radio3.ee.uec.ac.jp

***Abstract*—This study attempts an extension of previously proposed node pair selection scheme in multiuser Multiple Input Multiple Output (MIMO) communication network based on stable matching problem to the case in which one node can be a candidate of plural pair nodes. To cope with such a situation, repeated application of stable matching algorithm is considered, where total pairs are determined through some stages, each of which selects subset of pair nodes. Thanks to the polynomial complexity of matching algorithm, the proposed procedure still keeps low computational cost. Applications considered here are multiple point to multiple point multiuser system and the simultaneous selection of source-relay-destination pairs. Computer simulations are carried out to investigate the possibility of the effectiveness of the proposed concept. It has been shown that the proposed method is useful for the efficient node pair selection in multiple point system where one source node can transact plural destinations.**

***Keywords*–Node selection; Multiple Input Multiple Output (MIMO) network; multiuser; interference channel; stable matching.**

I. INTRODUCTION

Recently, MIMO interference channel [1] got attention since its utilization is widely found in many applications in communication networks [2]-[4]. Different from conventional multiuser MIMO, this system consists of plural nodes, as the source and destination, and the total performance depends on the node pair selection. The best way is the exhaustive search of the pairs with the best performance (since the number of node pairs is limited, the best solution could be always found at the expense of heavy computational burden), but it normally requires a considerable computation. On the contrary, fixed pair selection needs no computation and sometimes it could be a reasonable choice; but, under the channel variation, performance of this scheme is degraded.

In a previous work [5], we have presented a node pair selection scheme in relay-aided multi-point MIMO communication system based on stable matching theory (so called "stable marriage problem"). The preference of target node is first determined based on some kinds of metrics like the maximum eigenvalue, norm; it has been shown that the solution using Gale-Shapley algorithm [6] can improve the performance compared with the fixed pair case with low computational cost. This method is putting the importance on the simpleness of the procedure; the node pair selection is simply applied to each of two hops separately. But, we still have many applications in the area of communication engineering in which the conventional approach cannot be directly utilized. The typical example is the case where the source nodes can communicate with plural destinations which is found in multi-cellular communication. As one of such cases, this study attempts the extension of node pair selection scheme presented in [5] to the case in which one candidate can conform node pair with plural pair nodes. For this aim, repeated application of the stable matching algorithm is considered, where total pairs are determined through some stages, each of which selects subset of pair nodes. Here, we inspect the possibility of such algorithms through two types of problems including the above-mentioned typical example of MIMO interference channel.

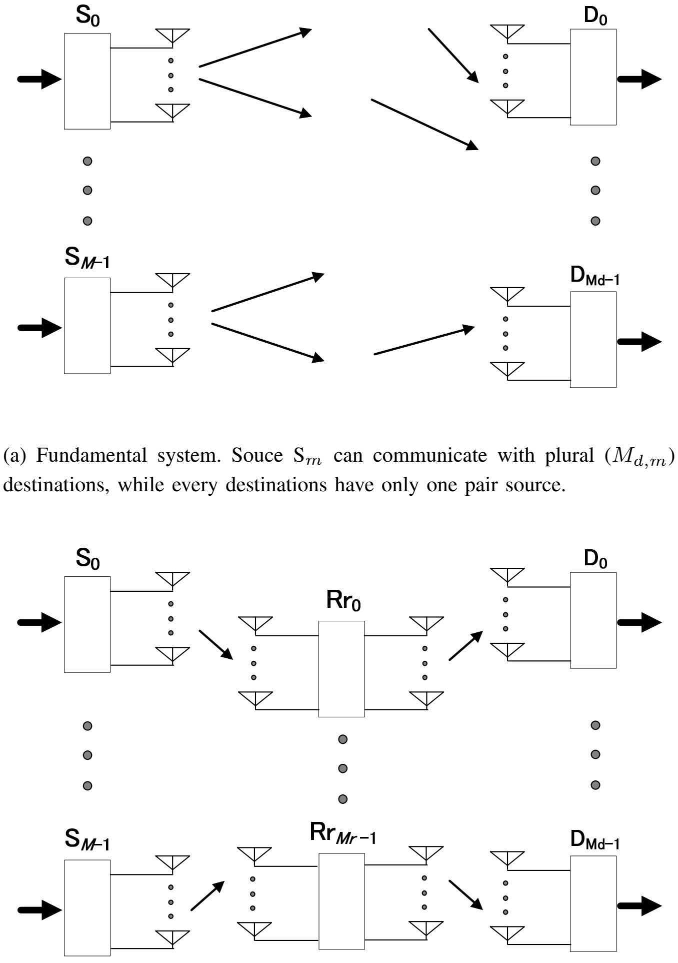
The organization of the rest of this paper is as follows: in Section II, the system model of two types of communication scenario is shortly given. In Section III, the detail of the proposed node pair selection scheme is described. Section IV evaluates the possibility of the effectiveness of the proposed approach through computer simulations. Finally, Section V provides the concluding remarks of this study and the future works.

II. SYSTEM MODEL

This section describes the two models of communication systems considered in this work.

A. Fundamental System

This fundamental system has the typical structure of multiple point to multiple point multiuser system, as shown in Figure 1 (a), consisting of M-sources ($\{\mathrm{S}_m;\ m = 0, \cdots .M-1\}$, S_m is equipped with $N_{s,m}$ antennas) and M_d-destination ($\{\mathrm{D}_{m_d};\ m_d = 0, \cdots .M_d-1\}$, D_{m_d} has $N_{d,m}$ antennas) nodes (this study assumes $M \leq M_d$). The channel between S_m and D_{m_d} is represented by $N_{d,m}$-by-$N_{s,m}$ matrix $H_{d,m_d,s,m}$ whose (n_d, n_s)-th element is a complex-valued response between n_s-th and n_d-th antennas of source and destination. Source S_m transmits L_m data streams $\{s_{m,\ell}(t);\ \ell = 0, \cdots, L_m - 1\}$ to destination D_{d_m} ($d_m \in \mathbb{Z}[0, M_d-1]$) which forms node pair with S_m through MIMO channel $H_{d,d_m,s,m}$, where weight vector $\boldsymbol{w}_{s,m,\ell}$ with $N_{s,m}$-dimension is used for the transmission of the ℓ-th data stream $s_{m,\ell}(t)$, and weight vector $\boldsymbol{w}_{d,d_m,\ell}$ with N_{d,d_m}-dimension is used for the production of ourput $\hat{s}_{m,\ell}(t)$. If the relation of source and destination is one-to-one, stable matching procedure can be directly applied for the determination of the node pairs. But here, we assume that S_m can transact plural destinations ($\mathrm{D}_{d_m,0}, \cdots, \mathrm{D}_{d_m,M_{d,m}-1}$). The modification in such a case is described in the following section.



(a) Fundamental system. Souce S_m can communicate with plural ($M_{d,m}$) destinations, while every destinations have only one pair source.

(b) Relay-aided system. Each of source, relay, and destination have one-to-one relation.

Figure 1. Model of multiple point to multiple point communication system.

B. Relay-Aided System

The construction of relay-aided system in this study is shown in Figure 1 (b). The difference from the previous model is in two points; (i) the system has Amplify and Forward (AF) relays $\{\mathrm{R}_{m_r};\ m_r = 0, \cdots, M_r-1\}$ with N_{r,m_r} antennas ($M \le M_r \le M_d$), and (ii) one-to-one connection among nodes (this is not mandatory, but here, we use this assumption to measure the effect of each scenario separately). Relay R_{r_m} receives signal from source S_m, produces output $s_{r,m,\ell}(t)$ (it is the estimate of $s_{m,\ell}(t)$ corrupted by noise) using weight vector $\boldsymbol{w}_{r,r,m,\ell}$ and retransmits it to destination D_{d_m} using weight vector $\boldsymbol{w}_{r,t,m,\ell}$. Hence, we need to determine M node pairs of source, relay, and destination; one method has been presented in [5] and another one is described in the next section.

C. System Design

In this study, weight vectors which appear in the two previous subsections are derived by channel inversion technique [7] (but it is applied to the product of the channel and destination weight, where destination weight is designed by singular value decomposition of the target channel as follows). Namely, the receiver vector of R_{r_m} detecting the ℓ-th data stream is derived as the right singular vector of $H_{x,m,y,n}$ $(((x,y) = (s,r)$ or $(r,d))$ for the fundamental system, and $((x,y) = (s,d)$ for the relay-aided system) corresponding to the ℓ-th largest singular value. Then transmit weight matrix is calculated by $\boldsymbol{w}_{x,m} = (H_{y,n,x,m}V_{0,m})^{-}$, where $((x,y) = (s,r)$ or $(r,d))$ or (s,d) and A^{-} denotes the pseudo-inverse of matrix A. The columns of $V_{0,m}$ spans kernel of $H_{y,-m,x,m}$, where $H_{y,-m,x,m}$ is given by removing $H_{y,m,x,m}$ from $[H_{y,0,x,m},\ \cdots,\ H_{y,M-1,x,m}$. This design method is adopted because (i) the design method with the best performance is not the main topic of this study, and (ii) it can avoid the problem of energy allocation, and suitable for the evaluatin of the pure effect of node selection.

III. Node Pair Selection

This section describes node pair selection schemes corresponding to each of two models shown in Section II.

A. Fundamental System

The total procedure consists of four steps, namely, (i) calculation of preference function $f_{x,n,y,m}$, (ii) determination of preference order list $\mathcal{P}_{x,m}$ from $f_{x,n,y,m}$, (iii) node pair selection based on $\mathcal{P}_{x,m}$, and (iv) removal of nodes which already have maximum number of pair nodes. The difference from the conventional approach is in that those steps are repeated to cope with the support of plural target nodes. The detailed manipulations are as follows:

(i) First, preference function $f_{x,n,y,m} = f_{y,m,x,n}$, $(x, y = s, d,\ x \neq y)$ is defined as a metric which shows how node x prefers node y. In this study, it is calculated based on the channel condition, since it is considered as an adequate index of the internode connection.

Here, we consider five types of metrics in the below.

(a) Fixed: Node pairs are fixed (e.g., $d_m = m$), which is equivalent to the case without node pair selection.

(b) Largest Eigenvalue: Preference function is $f_{d,n,s,m} = \lambda_0$, where λ_0 is the largest eigenvalue of $H_{d,n,s,m}$.

(c) Approximated Sum Capacity: Preference function is $f_{d,n,s,m} = \sum \log_2{(1 + \lambda_k)}$, where λ_k is the k-th largest eigenvalue of $H_{d,n,s,m}$.

(d) Norm: Preference function is $f_{d,n,s,m} = \|H_{d,n,s,m}\|_F$ (approximation of sum capacity).

(e) Absolute Sum: Preference function is $f_{d,n,s,m} = \sum_{p,q} |H_{d,n,s,m,p,q}|$, where $A_{p,q}$ is the (p,q)-th element of matrix A (simple approximation of sum capacity).

Among those, (b), (c) are based on eigenanalysis, (d) needs multiplication, and (e) does not require either, hence the computational burden of (e) becomes much lower. □

(ii) Second, the preference order list $\mathcal{P}_{x,m} = \{p_{s,m,0} \succ \cdots \succ p_{s,m,N-1}\}$ $(N = M, M_d)$ is obtained, where $x \succ y$ denotes that x is preferred to y. The list is made simply by the relation if $f_{x,p,y,m} > f_{x,q,y,m}$, then $p_{s,m,p} \succ p_{s,m,q}$. □

(iii) Third, the Gale-Shapley algorithm (for the detail of this algorithm, see [5] or [6]) is utilized to determine the node pairs based on preference order list. It is assured that the algorithm converges to stable pairs within polynomial time. Then, if S_m is connected to a destination with the largest preference function, they are recognized as a pair. If not, this source node proceeds to next iteration to search for a better candidate of the pair node. □

(iv) Fourth, if the number of pair nodes for node x_m has reached its maximum (here, $M_{d,m}$ for S_m, and 1 for all destinations), then x_m is removed from active nodes (active node means a node which has pair nodes less than its maximum and further looking for its next pair) and the preference order list of all active nodes. Then the procedure goes back to step (i), and repeats steps until no active source node is left. □

The above procedure requires repetition of original algorithm, but still keeps the polynomial computational time. Together with the fact that the node pair selection itself demandes just changing list elements several times without any multiplication, once the metric is given, the computational cost is very low.

B. Relay-Aided System Design

In the previous work [5], we have considered separate node pair selection in each of source-relay and relay-destination links, but another idea is to choose both link simultaneously. This scheme can be achieved by a similar procedure, as shown in Section III-A; this is the reason we consider this application here. The aim dealing with relay-aided system is to inspect which is better choice. In addition, the above idea has a possibility to be extended to multihop multinode systems, where method of [5] is not used; since it can include different number of hops depending on the routing. But here, we concentrate on the inspection of the effect in one-to-one relaying application.

The pair selection steps are shown below (for the discrimination from the step number (i), (ii), etc., in Subsection III-A, here, we use (1), (2), etc.).

(1) Consider a set of channels $\{H_{d,n,r,k,s,m} = H_{d,n,r,k}H_{r,k,s,m}\}$ and define preference function $f_{x,n,y,m,k} = f_{y,m,k,x,n}$, which can be regarded as a fundamental system of MM_r source nodes and M_d destinations. □

(2) Preference order list $\mathcal{P}_{x,m} = \{p_{y,m,0} \succ \cdots \succ p_{y,m,N-1}\}$ $(N = MM_r, M_d)$ is made, and node pair selection is carried out in the same manner as step (ii) of the previous section. □

(3) If the pairs of different destinations contain *same source or relay*, then the destination with larger preference function is chosen. After that, the source, relay, and destinations which already have pair nodes are removed from active nodes. The elements concerning those removed nodes are concurrently removed from preference list of all active nodes, and steps (1)∼(3) are repeated until no active source node is left. □

This procedure is assured to converge since that of each iteration is guaranteed. The comparison with the conventional approach is made in the next section.

IV. Computer Simulation

Default simulation conditions are enumerated in Table I. In this section, computer simulations are carried out to investigate the effectiveness of the procedures described in Section III.

TABLE I. Default Simulation Conditions.

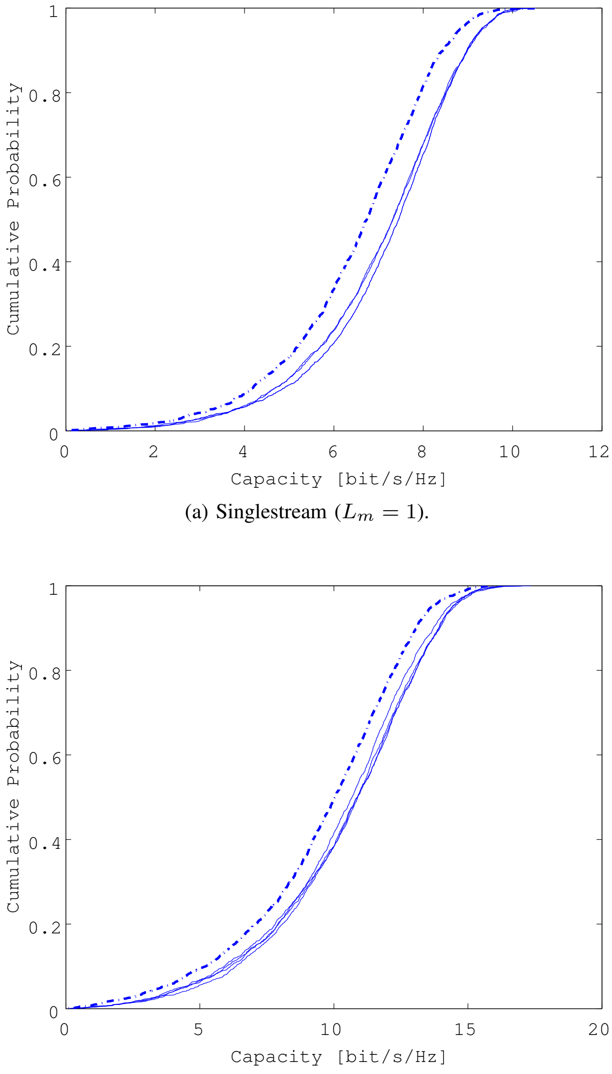
User Number	$M = 3$
Node Number	Source $M = 3$ Relay $M_r = 5$ Destination $M_d = 8$
Antenna Number	Source $N_{s,m} = \begin{cases} 6 & (L_m = 1) \\ 12 & (L_m = 2) \end{cases}$ Relay $N_{r,m} = 6$ Destination $N_{d,m} = 2$
Data Stream Number	$L_m = 2$
Modulation	QPSK
Relaying Scheme	Amplify and Forward (AF) half-duplex
Relay SNR	SNR$_{r,m} = 5 \sim 30$ dB (default : 20dB)
Destination SNR	SNR$_{d,m} = 5 \sim 30$ dB (default : 20dB)
Fading (fundamental)	$H_{d,n,s,m}$ i.i.d. Quasistatic Rayleigh
Fading (relay-aided)	$H_{d,n,s,m}$ ignored $\left.\begin{array}{l} H_{r,n,s,m} \\ H_{d,n,r,m} \end{array}\right\}$ i.i.d. Quasistatic Rayleigh

The evaluation is by sum capacity represented by $C_m = \sum_{\ell} \log_2(1 + \gamma_{m,\ell})$ for the m-th user, where $\gamma_{m,\ell}$ is the Signal to Interference plus Noise Ratio (SINR) of the ℓ-th data stream of the m-th user, which should be discriminated from Signal to Noise Ratio (SNR) SNR$_{x,m}$ $(x = r, d)$ defined as energy ratio of transmitted signal and the receiver noise. The modulation scheme is QPSK (a fundamental scheme with PAM in both of real and imaginary axes), but the effectiveness of the proposed scheme is not affected by the choice of modulation.

In the fundamental system (Section III-A), the number of users is $M = 3$, while that of destination is $M_d = 8$, where all of them are equipped with $N_{s,m} = 6$ (singlestream) or $N_{s,m} = 12$ (multistream) and $N_{d,m} = 2$ antennas, respectively. The number of data streams is $L_m = 1$ or 2, and one source node supports two users, hence $M_{d,m} = 2$. In the relay-aided system (Section II-B), where $N_{r,m} = 6$ and other conditions are same as multistream transmission in the fundamental case, the first hop is assumed not Ricean (LOS: Line Of Sight), but Rayleigh fading with unit variance since (i) there are applications this assumption applies [8], and (ii) in case of LOS, it is clear that the effect of node pair selection in the first hop is small when direct path is dominant.

Under those simulation conditions, the mean statistics are calculated using 2,000 samples of fading channels.

Figure 2 shows empirical distribution functions of sum capacity in fundamental system. Here, because of the symmetry of the channels among nodes, the curves of mean value over $\sum M_{d,m}$ users are considered. In both subplots for (a) singlestream ($L_m = 1$) and (b) multistream ($L_m = 2$), the curves of node pair selection (four solid lines corresponding to the metrics (b)sim(e)) overcome that of fixed case (broken line corresponding to the metric (a)), which shows the proposed approach in Section II-A has certain effectiveness. (Since the curves of (b)∼(e) overlap and not identifiable, all of them are written by solid line. For the comparison of (b)∼(e), see Figure 3.) The amount of improvement seems not so large, but we should remark that this is achieved only by the exchange of node pairs, and in particular adopting absolute sum preference function, no multiplications are required. In both figures, one of preference functions (in (a) absolute sum,



(a) Singlestream ($L_m = 1$).

(b) Multistream ($L_m = 2$).

Figure 2. Distribution function of capacity in fundamental system shown in Figure 1 (a).

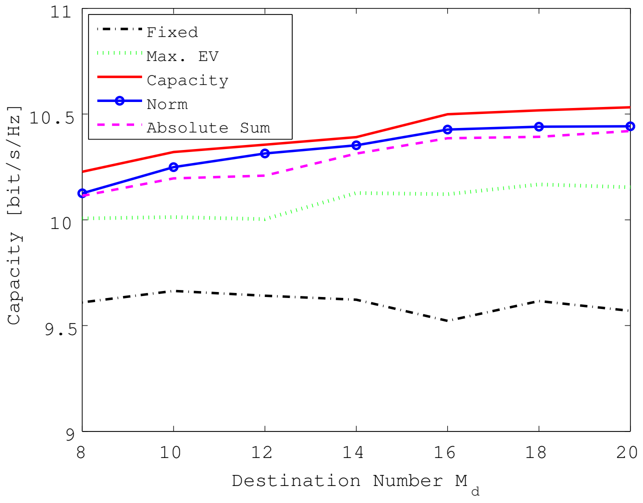


Figure 3. Destination number versus capacity in fundamental system ($L_m = 2$).

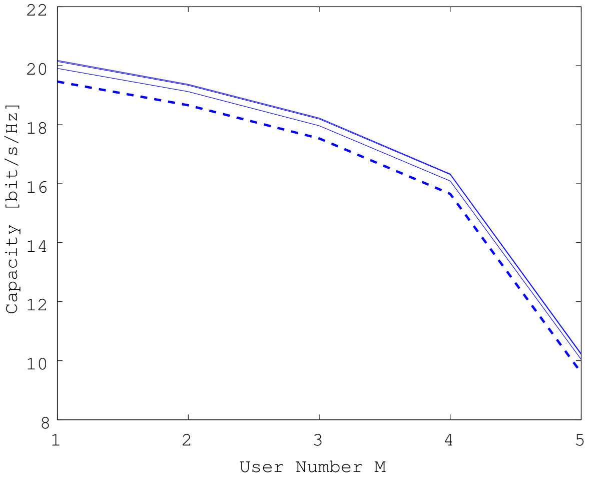


Figure 4. User number versus capacity in fundamental system ($Ns, m = 20,\ L_m = 2$).

and in (b) maximum eigenvalue) is inferior to others, but only quite small difference can be observed.

The difference among five preference functions more clearly can be seen in Figure 3, which depicts the relation between the destination number M_d versus capacity in multistream case. Though small perturbation by the randomness of the sample appears, we can observe the trend of the curves; as M_d increases, the effect of node pair selection becomes larger. From this figure, it can be seen that the performance of capacity improvement is the largest using capacity preference function, but absolute sum without multiplication attains capacity close to it, which is good nature from the viewpoint of computational cost. In this multistream case, the maximum eigenvalue is not a suitable function.

The relation between user number $M = 1 \sim 5$ and capacity (per user) is depicted in Figure 4. In this figure, every source again supports two users ($N_{d,m} = 2$), hence each of them should deal with $N_{d,m}L_m$ streams (for example, in case of $M = 5$ and $L_m = 2$,, totally 20 streams). Therefore, the number of source antenna is changed to $N_{s,m} = 20$. We can observe from the graph that the capacity decreases as the user number increase, but what is important is the effectiveness of the node pair selection is not decreased.

Turning to the target to the relay-aided system, distribution functions of capacity for fixed (broken line, preference function (a) in Section III) and node pair selection (two solid lines) adopting absolute sum preference function ((e) in Section III) are drawn in Figure 5. The latter represented by two solid curves correspond to two methods, namely, separate selection in two hops [5] and the approach shown in Section III-B. From this figure, what can be observed is they have quite similar characteristics. Therefore, in this situation, the proposed method does not have advantage against the previous one in [5] which has simpler procedure, so we need to search another application in multihop network where method [5] cannot be directly used.

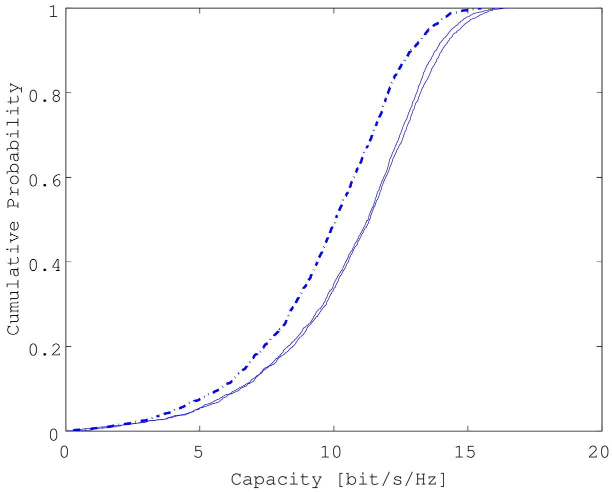


Figure 5. Distribution function of capacity in relay-aided system shown in Figure 1 (b) ($L_m = 2$).

V. Conclusion and Future Work

This study has presented the extension of node pair selection scheme for multiuser MIMO communication based on stable matching problem to the case where one node can be a candidate of plural pair nodes. The repeated application of stable matching algorithm with the polynomial complexity is considered, which consists of some stages in each of which a subset of pair nodes is determined. Computer simulations have been carried out to investigate the possibility of the effectiveness of the proposed concept. The proposed method is useful for the node pair selection in multiple point system in which one source node can transact plural destinations.

A future work is the extension of the proposed method to efficient multihop rooting in multinode communication system, where the paths with different number of hops become candidates of the selection.

Acknowledgment

This research was supported in part by the Grant-in-Aid for Scientific Research (No. 25249056) from Japan Society for the Promotion of Science (JSPS). Authors will thank to Dr. Nobuo Nakajima, UEC, for his helpful discussion about the initial idea of this work.

References

[1] M. Maddah-Ali, A. Motahari, and A. Khandani, “Communication over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis,” IEEE Trans. Inf. Theory, vol. 54, no. 8, Aug. 2008, pp. 3457-3470.

[2] T. Q. S. Quek, G. de la Roche, I. Gueven, and M. Kountouris (Eds)., Small Cell Networks: Deployment, PHY Techniques, and Resource Management, Cambridge University Press, Cambridge, UK, 2013.

[3] F. Monsees, C. Bockelmann, D. Wubben, and A. Dekorsy, “A sparsity aware multiuser detection for machine to machine communication,” 2012 IEEE Global Commun. Conf. Workshops (Globecom Workshops), Dec. 2012.

[4] M. Barbeau and E. Kranakis, Principles of Ad-hoc Networking, Wiley, Sussex, UK, 2007.

[5] T. Taniguchi and Y. Karasawa, “An elementary study on node pair selection in relay-aided communication system based on stable marriage problem,” Tech. Rep. IEICE, RCS2014-50, June. 2014.

[6] A. E. Roth and M. A. O. Sotomayor, Two-sided matching: A study in game-theoretic modeling and analysis, Cambridge University Press, Cambridge, UK, 1990.

[7] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, “An introduction to the multi-user MIMO downlink,” IEEE Commun. Mag., vol. 42, no. 10, Oct. 2004, pp. 60-67.

[8] S. Loyka and G. Levin, “On outage probability and diversity-multiplexing tradeoff in MIMO relay channels,” IEEE Trans. Commun., vol. 59, no. 6, pp. 1731-1741, June 2011.

Evaluation of the New e-Health Signaling Model in the Ubiquitous Sensor Network Environment

Adel Mounir Said
Faculty of Engineering
Ain Shams University
Cairo, Egypt.
amounir@nti.sci.eg

Ashraf William Ibrahim
Switching Department
National Telecommunication Institute – NTI
Cairo, Egypt.
awilliam@nti.sci.eg

***Abstract*— Ubiquitous Sensor Network (USN) is a conceptual network built over existing physical networks. It makes use of sensed data and provides knowledge services to anyone, anywhere and at anytime, and where the information is generated by using context awareness. In 2010, the ITU-T provided the requirements to support USN applications and services in the Next Generation Network (NGN) environment to exploit the advantages of the core network. One of the main promising markets for the USN application and services is the e-Health. It provides continuous patients' monitoring and enables a great improvement in medical services. In this paper, the authors provide the evaluation of the e-Health signaling model in the USN environment, which was introduced in a previously published work. The model is based on using the IP Multimedia Subsystem (IMS) as a service controller sub-layer for the USN platform. This paper provides a USN based IMS detailed network design for e-Health implementation with emphasizes on Session Initiation Protocol (SIP) modification and middleware entities functions. The proposal evaluation was carried using OPNET Modeler for network simulation and proved its applicability and reliability.**

Keywords- e-Health; ubiquitous sensor netowrk; NGN; IMS; SIP.

I. INTRODUCTION

The ITU recommendation Section for Ubiquitous Sensor Network (USN) presents the requirements for a platform to numerous number of life services and applications [1]. The USNs consist of collaborative efforts of many small wireless sensor nodes. These nodes are small and autonomous devices capable of measuring all sorts of environmental and physical conditions (e.g., temperature, sound, vibration, pressure, motion or pollutants) forming the USN sensor layer. The Next Generation Network (NGN) capabilities can impact the service requirements of the USN. Support of USN applications and services may require some extensions and/or additions to NGN core.

One of the main emerging markets for the USN applications and services is the e-Health. It has been invented to exploit the wide use of the USNs to gather the patients' medical/non medical data for various applications. e-Health offers timely coverage of the advances in technology that offer new and innovative options to practitioners, medical centers, and hospitals for managing patient care, electronic records, and many other features.

Considerable ongoing research efforts are focusing on providing the physical design for the USN entities according to the ITU requirements. Some of these proposals are based on Context Awareness (CA), such as [2][3][4]. Also, an integration between service enablers based on the CA system is proposed to provide various services in [5]. Paganelli et al. proposed a context aware mobile service platform supporting mobile caregivers in their daily activities [6]. They have demonstrated its capability for providing an extensible set of services aiming at supporting care networks in cooperating and sharing information for the goal of improving a chronic patient's quality of life. This work was used for health monitoring and alarm management of chronic conditions in a home-based care scenario [7]. The CA service architecture proposed by Domingo [8] can also be used in order to integrate with social networks.

Another group of researches concentrated on providing a network and service integration techniques for the USN with the IP Multimedia Subsystem (IMS). One of the recent solutions is based on building a service enabler over the IMS to support e-Health services without mentioning the consideration of the core network [9][10].

Most of these researches and standardization efforts proposed only theoretical ideas and potential scenarios. However, they did not provide a detailed signaling flow to be the roadmap for implementing these e-Services.

Apart from the ITU recommendations of USN, there are other approaches for building a platform for sensor networks services. Singh et al. [11] proposed a prototype for a global homecare monitoring system. The prototype is based on using IP-based USN in a personal area to provide sensor data. Kim [12] proposed a middleware platform including several types of sensor network for building USN services. The platform is used to build a healthcare service that is proposed to be integrated with standardized medical devices communication framework based on ISO11073/ IEEE1073 standard. These proposals introduce a healthcare service with minor capabilities offering monitoring only due to the limited core network facilities. Moreover, it did not consider the challenges of the integration between USN and existing infrastructure. Besides that, these solutions miss proposing real scenarios and its required signaling flow.

On the contrary, in [13], we proposed the usage of IMS as a service controller sub-layer in the USN environment. The IMS platform is used to utilize its benefits and features

ISBN: 978-1-61208-368-1

[9] and to provide the service requirements of USN applications and services [1]. The main contribution of [13] is to develop a detailed network signaling flow for different applicable e-Health scenarios using Session Initiation Protocol (SIP).

In this paper, we provide a USN-based IMS detailed network design for implementing the e-Health service with emphasizes on the middleware layer entities functions. There is a need to modify the SIP protocol (SIP MESSAGE request) to match the features provided in the proposed e-Health service as will be discussed in details in Section III.

The rest of the paper is organized as follows: Section 2 describes the detailed proposed network architecture. Section three explains the modification of the SIP Message Request and initialization phase, as well as the different applicable scenarios. Section 4 presents the proposal evaluation using OPNET simulation. Finally, Section 5 gives a conclusion and an idea about the current and future work.

II. PROPOSED NETWORK AND SERVICE ARCHITECTURE

The proposed network follows the IMS-based NGN architecture and according to the requirements of the USN [1]. We implemented these requirements in the form of physical devices as described in the proposed model as shown in Figure 1. The architecture is divided into two stratums: transport stratum and service stratum.

The patient sensor network contains different types of sensor nodes with wireless capability. These nodes observe the patient and the surrounding atmosphere as well. A USN Gateway is used to translate between the network's access network protocol and that of the sensor nodes providing the connectivity requirement to the IMS infrastructure.

The transport stratum includes transport sub-layer and transport control sub-layer. The transport sub-layer contains access network and core transport network. The access network take care of end-users' access to the network as well as collecting and aggregating the traffic towards the core network. The transport control sub-layer is further divided in two subsystems the Network Attachment Subsystem (NASS) and the Resource and Admission Control Subsystem (RACS) to provide the QoS, privacy, security, and authorization, as required for the USN applications. These elements provide transport control functionalities according to the standards [14][15].

The service stratum includes IMS service control sub-layer, USN middleware sub-layer, and the IMS service application sub-layer.

It provides the platform for enabling services to the user. It includes registration and session control functions. It includes three sub-layers: the service control sub-layer, the USN middleware sub-layer, and the service application sub-layer. The service control sub-layer is based on the standard IMS [16] and controls the authentication, routing, and database of the subscribers. This sub-layer provides the USN requirements needed including service profile, open service environment, security, and authorization.

The USN middleware sub-layer consists of a set of logical functions to support USN applications and services. It contains the Application Servers (ASs) providing the different services. It interfaces with the Serving Call Session Control Function (S-CSCF) using SIP and is responsible of applications execution.

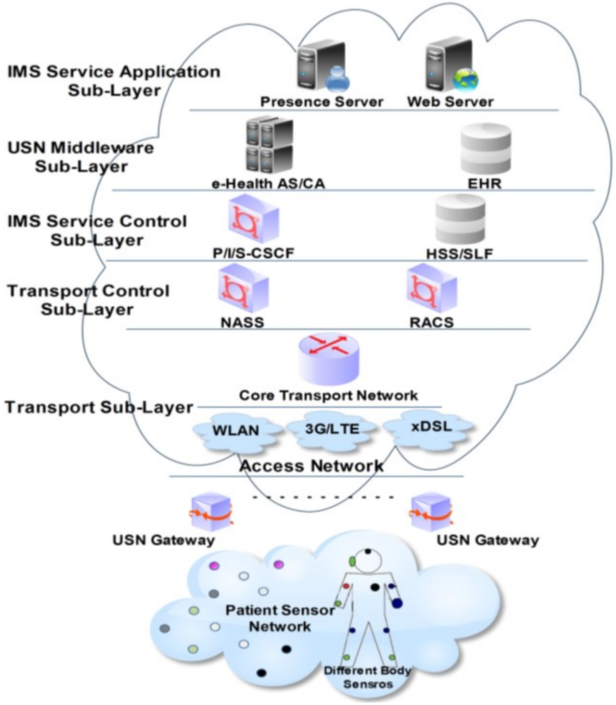


Figure 1. Next Generation e-Health Network Architecture

There is a need for a CA server to automatically adapt an application or service depending on the user current situation. In this proposal, a new AS integrated with a CA server is developed to provide e-Health services. This integration eases services control and reduces the signaling required between both of them. There is one centralized server, which provides a coherent environment and is responsible of services control.

The USN middleware sub-layer contains also a private database for the e-Health services’ subscribers. We call it the Electronic Health Record (EHR). It contains the initial sensors configuration settings, the collected monitored vital signs, the different patients’ data files such as (X-Ray, tests results, prescription, etc.), emergency contacts, medical supervisors, medical history, and any other information related to the patient health. The proposal allows the patients to access the EHR. This enables them to change their details, update their emergency contacts, and upload files. Access to the EHR is provided through a Web Server (WS), which is implemented in the service application sub-layer or via an IMS client.

This sub-layer covers the sensor network management, service profile, open service environment, location based service support, and service privacy of the USN requirements.

The Service Application Sub-layer contains a Presence Server (PS), which is used to follow and publish the patients’ status in real-time to their emergency contacts as explained later in details. This addition allows informing selected persons by the emergency situation as soon as it occurs.

This sub-layer contains also a Web Server (WS), which interfaces the patients' records. Through the WS, patients can change in their details, upload files, contact medical centers, doctors, etc.

III. Signaling Scenarios

This section focuses on the IMS and SIP functionalities and does not cover the transport layer as there is no modification to its functions.

The Section describes in details the initialization phase as well as the different applicable scenarios. It is to be noted that the SIP "MESSAGE" requests in these scenarios are used differently from their original use. The standard SIP "MESSAGE" requests are designed to carry content in the form of Multipurpose Internet Mail Extensions (MIME) body parts. Therefore, in this work, we propose using a special format and data fields for the MESSAGE request body to carry data from/to AS/CA. These data may contain sensors initial configurations, sensors data, alerts, subscribers' information, etc. Figure 2 shows an example of the proposed data fields provided in the MESSAGE request body.

A. Initialization and Registration

Figure 3 shows the registration messages for a patient, which follows the standard IMS client registration until message 20. The messages flow starting from message 21 until the end message presents the proposed signaling flow to initialize and activate the e-health service and the modification of the SIP MESSAGE to fulfill the requirements of the service. After the registration completion, the S-CSCF evaluates the user's Initial Filter Criteria (IFC) (message 21) and accordingly, it forwards the register message to the AS/CA (22). Based on the e-heath algorithm saved in the AS/CA, it downloads the initial configurations, measurements thresholds, possible diseases' situations, and patient's emergency contact list from the EHR using the HTTP "GET" request (24). HTTP is proposed to be used between the EHR/WS and the AS/CA server. It is to be noted that every patient (e-Health subscriber), via the WS interface, can build his own contact list to be notified in case of emergency as mentioned before. The AS/CA forwards the downloaded initial configurations to the USN Gateway in the body of a SIP "MESSAGE" request (26). The next message is that the AS/CA sends a "SUBSCRIBE" request (32) to the PS to be notified by the presence status of the persons in the downloaded patient's emergency list. The PS sends back a "NOTIFY" message (34) containing the current status. The PS sends a "NOTIFY" request to the AS/CA every time there is a status change. At the same time, the AS/CA sends a "PUBLISH" request (36) to the PS containing the current patient's status. It is to be noted here that the patient's status is not meant to be online or offline. Instead, it reflects the patient's health condition (normal, critical, emergency, etc). Another important issue arises; the contacts in the patient's emergency list (MC: Medical Center, Rel: Relatives, Doctors, etc.) have to follow his status automatically and without their intervention. To solve this, the AS/CA subscribes on behalf of them to the patient's status by sending a "SUBSCRIBE" request (38) to the PS. Accordingly, they will be continuously notified of any status change via "NOTIFY" requests (40).

Request Line
Header
.
Message Body

Fields of: Sensors Initial Configuration
OR
Fields of: Subscriber Information
OR
Fields of: Sensors Data
OR
Fields of: Alerts Types & Severity Degree

Figure 2. Proposed SIP Message Request Body for e-Health Services

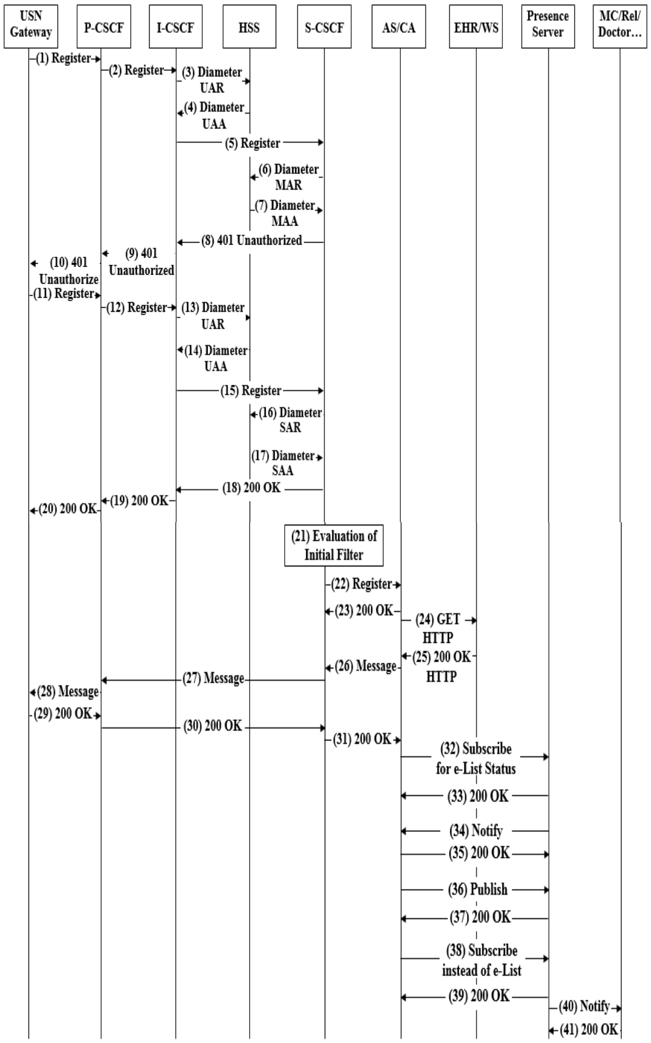


Figure 3. Registration and Initialization Scenario

B. *First Scenario (Periodic Transmissions)*

This Scenario proposes a simple case for transmitting periodic sensed information. The period to transmit regular collected data is determined from a timer value in the initial configuration received from the AS/CA during the initialization phase. The USN Gateway collects the sensed data from the sensors. When the data is ready for transmission, the USN Gateway puts the collected sensors data in the body part of a SIP "MESSAGE" request (1) and sends it to the AS/CA. the request is routed normally to the AS/CA, which stores the data in the EHR using HTTP "PUT" (4). Figure 4 shows the scenario's signaling flow. The USN Gateway transmits the data to the EHR through the AS/CA and not directly, because the collected data has to be assessed and compared to specific thresholds to determine emergency cases (more details in next scenario). These thresholds are set in the initial configuration file downloaded from the EHR during the initialization. This task has to be done by the AS/CA and not the EHR as this later is only a database, and it has no control or service algorithm as in the AS/CA. Another reason is security as the EHR must be hidden from non-trusted users' devices.

C. *Second Scenario (Emergency Case)*

Figure 5 shows the emergency scenario. The emergency case is determined if the collected data values are out of the threshold range set in the initial patient's file stored in the EHR and downloaded to the AS/CA. After saving the users' data into the EHR, the AS/CA evaluates the collected values to the preset thresholds. If an emergency case is identified, it updates the patient status in the PS using a "PUBLISH" request (10) to be critical or emergency. The PS updates, in turn, the emergency status in the patient's contacts list by a "NOTIFY" request (12). Simultaneously, an alert message is sent to these contacts. The AS/CA has two options to do so depending on the contacts' IMS status: online or offline. In case the contact person is online, the alert is sent using a SIP "MESSAGE" (14) containing the current patient's data. In case the contact person is offline, the alert is sent using a SMS message (16) through the mobile network. The contacts' IMS status is known since the AS/CA has already subscribed to their status during the initialization phase. The emergency contacts could be relatives, treating doctor, medical center, ambulance, neighbors, etc.

D. *Third Scenario*

This scenario, shown in Figure 6, provides the case of a patient uploading a file (scanned X-ray, ultrasound, magnetic resonance, etc.) to his EHR record. There are two options for doing this. The first option uses an IMS client. When the uploader (patient) needs to upload a file, he transmits a SIP "MESSAGE" (1) containing in its body part a request for the necessary upload information (URL, username, password, etc.). Once the request arrives at the AS/CA, it asks the needed information from the EHR by a HTTP "GET" (7). The AS/CA forwards this information to the uploader in a SIP "MESSAGE" (9). The uploader can now upload the file according to the received settings using FTP (15). The username and password sent in the previous message are temporary and will change the next time for security. After uploading the file successfully, the EHR informs the AS/CA of the upload termination using HTTP "POST" (17). According to the patient's customized service algorithm, the AS/CA informs the concerned persons (e.g., doctor, medical center, relatives, etc.). This is done, depending on the contacts' IMS status, using a SIP "MESSAGE" request (19) or a SMS (21) as explained in the previous scenario.

The second option is shown in Figure 7. In that case, the uploader does not use an IMS client. He uses HTTP to browse the WS (1), and submits his file directly by FTP (2, 3) without using the IMS network. As in the first option, the EHR informs the AS/CA of the upload termination using HTTP "POST" (4). The AS/CA, in turn, informs the concerned persons using a SIP "MESSAGE" request (6) or a SMS (8).

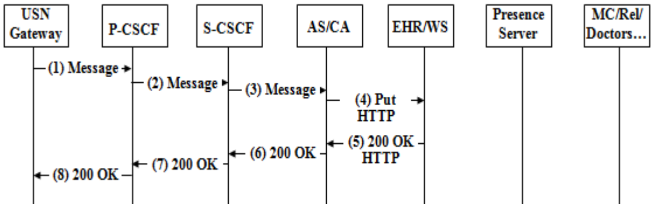


Figure 4. Periodic Transmission Scenario

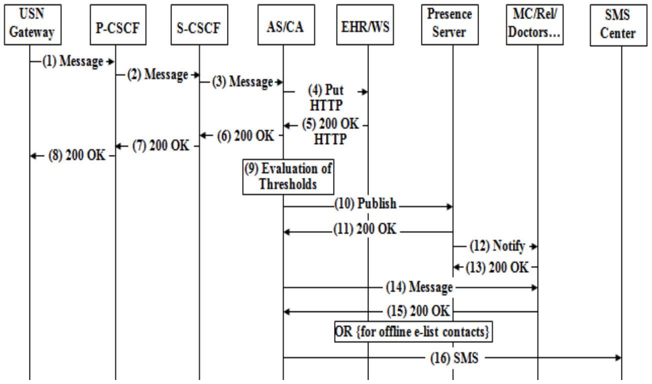


Figure 5. Emergency Scenario

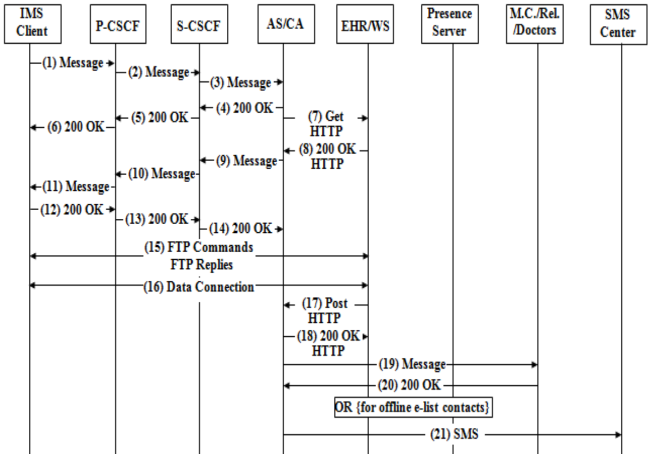


Figure 6. Uploading Patients' Files using an IMS Client

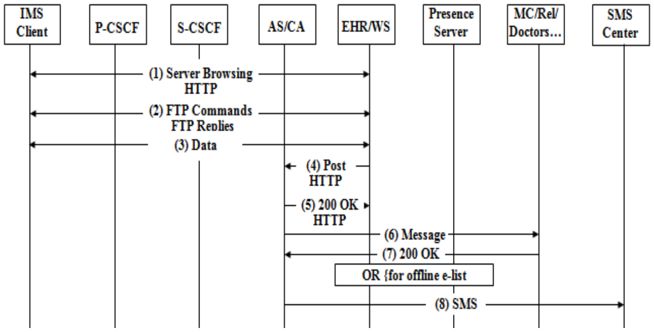


Figure 7. Uploading Patients' Files Through the Web Server

E. *Fourth Scenario*

This scenario is similar to the previous one. However, the uploader here is not the patient but the treating doctor or the scanning center or the tests laboratory, etc. Consider the patient's doctor would like to send or update the prescription of the patient or the lab wants to send the test results. They will connect directly to the WS using their accounts credentials and upload the new file or update the existing information. The sender can also identify the urgency of this data. The EHR informs the AS/CA of the new upload or the information change. The AS/CA, in turn, informs the patient using a SIP "MES-SAGE" request or a SMS depending on his IMS availability.

IV. PERFORMANCE EVALUATION

This section presents the performance evaluation of the proposal. To evaluate the system, we implemented the different scenarios using the OPNET modeler 14.0. The objective of the simulation is to study the applicability and reliability of the proposal under different conditions such as number of users and links bandwidths (BWs). As this work concentrates on the IMS and the SIP protocol and as the NGN transport layer has no change in our model, the performance Section focuses on the IMS evaluation. The simulation model passes by sequential steps as follows:

- Build the proposed network architecture (IMS servers) using the OPNET simulation tool as shown in Figure 8 and assign the transmission delay to the servers' connections according to the values of Table I.

- Create the developed signaling flow using the graphical OPNET application task tool (ACE whiteboard) as shown in Figure 9. The servers are called tiers in ACE whiteboard.

- Assign the processing time of each server (tier) in the ACE whiteboard model according to Table I in [17].

- Import the ACE whiteboard model into the main project model.

- Assign the server function of each ACE model tier corresponding to the developed signaling flows.

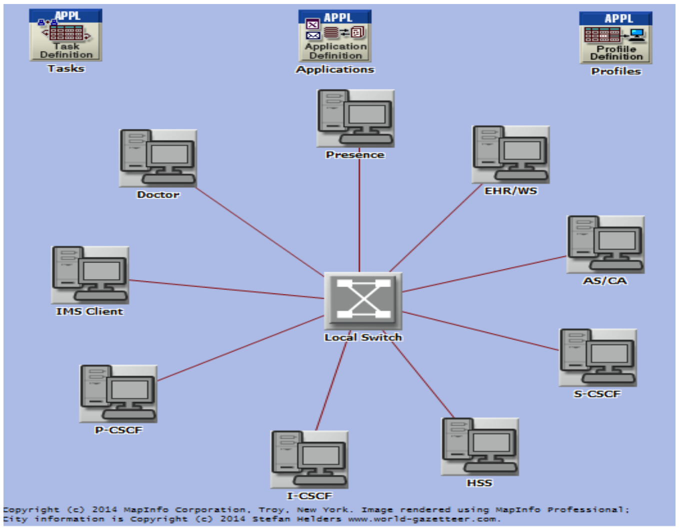
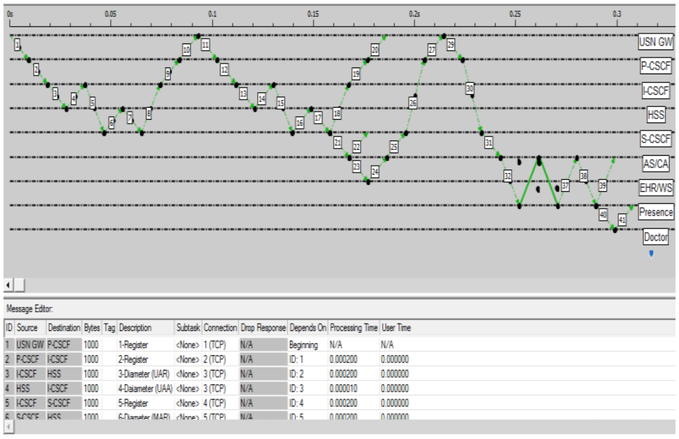

APPL
Task Definition
Tasks
APPL
Application Definition
Applications
APPL
Profile Definition
Profiles
Doctor
Presence
EHR/WS
AS/CA
IMS Client
Local Switch
S-CSCF
P-CSCF
I-CSCF
HSS
Copyright (c) 2014 MapInfo Corporation, Troy, New York. Image rendered using MapInfo Professional; City information is Copyright (c) 2014 Stefan Helders www.world-gazetteer.com.

Figure 8. Network Architecture Built using the OPNET Simulation Tool

TABLE I: IMS ENTITIES PROCESSING AND TRANSMISSION TIMES

Parameters		Duration
Process time (microsecond)	UE P/S/I-CSCF HSS	200 200 10
Transmission delay (microsecond)	UE/P-CSCF other links	5,000 200



Message Editor:

ID	Source	Destination	Bytes	Tag	Description	Subtask	Connection	Drop Response	Depends On	Processing Time	User Time
1	USN GW	P-CSCF	1000		1-Register	<None>	1 (TCP)	N/A	Beginning	N/A	N/A
2	P-CSCF	I-CSCF	1000		2-Register	<None>	2 (TCP)	N/A	ID: 1	0.000200	0.000000
3	I-CSCF	HSS	1000		3-Diameter (UAR)	<None>	3 (TCP)	N/A	ID: 2	0.000200	0.000000
4	HSS	I-CSCF	1000		4-Daiameter (UAA)	<None>	3 (TCP)	N/A	ID: 3	0.000010	0.000000
5	I-CSCF	S-CSCF	1000		5-Register	<None>	4 (TCP)	N/A	ID: 4	0.000200	0.000000
6	S-CSCF	HSS	1000		6-Diameter (MAR)	<None>	5 (TCP)	N/A	ID: 5	0.000200	0.000000

Figure 9. ACE Whiteboard Model for the Registration and Initialization Signaling Flow

Simulation 1: for the first scenario (initialization and registration), we studied the effect of increasing the BW of the link between the USN Gateway and the network on the initialization time. The initialization time is defined as the time spent between the register message (1) and the 200 OK message (41). We repeated the simulation several times increasing the BW from 50 to 500 Kbps in steps of 50 Kbps. The connection BW is chosen to be low to assure the reliability of the proposed scenarios.The initialization time was calculated for different number of users varying from 100 to 1000. Figure 10 shows the collected results. The simulation shows that the delay of the initialization time decreases with the increase of the connection BW. On the other hand, the number of users has a negligible effect for bandwidths higher than 150 kbps and a small effect for bandwidths less than 150 kbps. This is because the connection bitrate required for each user to register and activate his e-health subscription is low according to the signaling proposed in Figure 3.

Simulation 2: the emergency detection scenario is evaluated by simulating the signaling flow shown in Figure 5. The simulation was run for different USN Gateway connection BWs ranging from 50 to 500 Kbps for 100, 500, and 1000 users respectively. The collected results are shown in Figure 11. The simulation shows that the maximum detection time of an emergency case is 0.21 seconds. This time decreases exponentially as the BW increases and is not affected significantly by the number of users. As shown in the figure, the number of users has a minimum effect on the delay. This is because the BW needed for the process is low, which proves the reliability of the developed signaling flow. Moreover, the delay is within an acceptable margin for implementing the emergency scenarios as they do not require high connection bitrate.

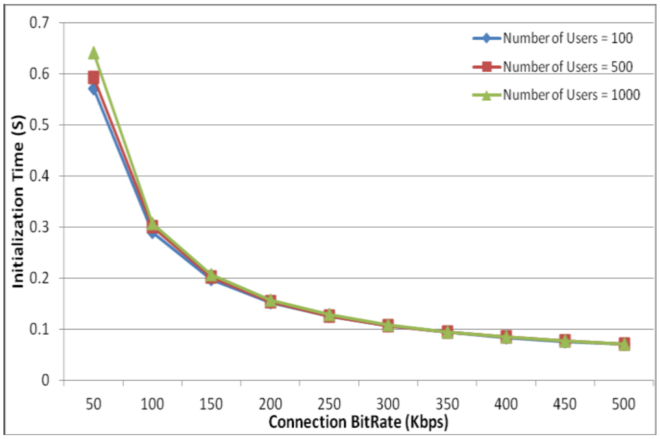


Figure 10. The Registration Time at Different Connection Speeds

TABLE II: UPLOADED FILES SIZES

File Type	Purpose	Average File Size
Text	Sensed Information Medical Analysis, etc.	A few KB
Image	X-Ray Magnetic resonance, etc.	0.5 MB
Video	Ultrasound, etc.	5.0 MB

TABLE III: HSUPA CATEGORIES SPEED

HSUPA Category	Max Uplink Speed
Category 1	0.73 Mbit/s
Category 2	1.46 Mbit/s
Category 3	1.46 Mbit/s
Category 4	2.0/2.93 Mbit/s
Category 5	2.0 Mbit/s
Category 6	2.0/5.76 Mbit/s

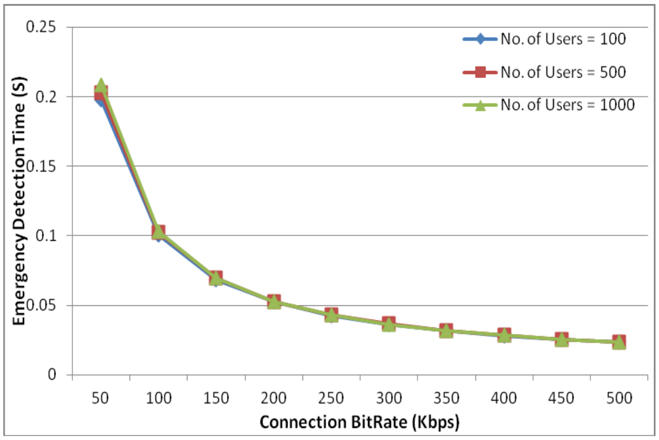


Figure 11. The Emergency Detection Time at Different Connection Speeds

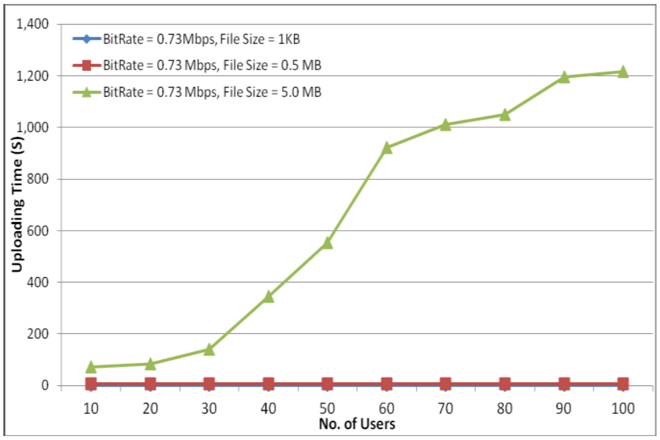


Figure 12. File Uploading Time at Different Number of Users

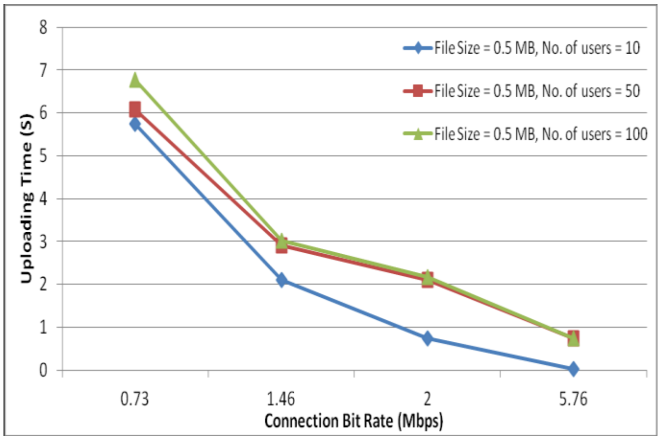


Figure 13. File Uploading Time at Different Connection Speeds

Simulation 3: to investigate the time required to upload a patient’s file. We simulated the flow shown in Figure 7 increasing the number of users from 10 to 100 in steps of 10 and the simulation was repeated for different file sizes as shown in Table II. The BW of the link between the USN Gateway and the network was set to 0.73 Mbps corresponding to HSUPA-category 1 connection rate. Figure

12 shows the output results. The figure clearly shows that at small file sizes of 1KB and moderate file sizes of 0.5 MB, the delay is almost constant with the number of users. These files sizes represent the USN sensed information or patient analysis files and the average size of the x-ray or image files respectively. On the contrary, for large file sizes (5.0 MB), which represent an average video file size, the delay increases with the number of users. Hence, the proposed solution is very effective in case of small or medium files representing data and radiography files. On the other hand, video files will suffer of delay depending on their sizes.

Simulation 4: for the same scenario of Figure 7, the effect of the USN Gateway connection BW on the upload time is studied. The simulation was run several times for different connection bit rates corresponding to the different HSUPA categories as stated in Table III. It is to be noted that the HSUPA Category 4 speed is not simulated as it is not being widely implemented. The file size was assumed to be 0.5 MB in this simulation. Figure 13 shows the results. The simulation proves that for a fixed number of users and file size, the delay decreases with the increase of the users' connection speed.

Simulation 5: the initialization process of Figure 3 and the upload scenario of Figure 7 were combined together in one final simulation to study their effect on each other. We simulated 10 users performing the initialization process simultaneously with 10 other users uploading files of 0.5 MB. The USN Gateway connection BW is 1.46 MB/S (HSUBA Category 2). The simulation was repeated for 50 and 100 users. The results are shown in Figure 14 along with the results of the initialization and upload times collected from the previous simulations where the two processes were simulated separately. The results show a slight effect on the upload time, which could be neglected. On the other hand, there is no congestion effect on the users registering to the network even when there are other users uploading files. This is because the registration process does not mandate high connection speed.

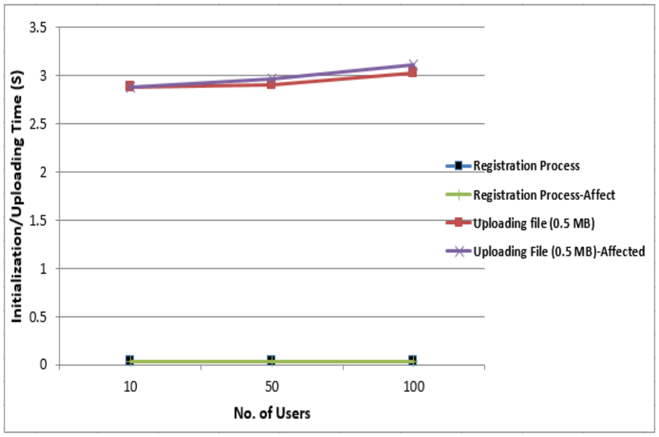


Figure 14. The Effect of Uploading FilesScenario on the Registration Scenario.

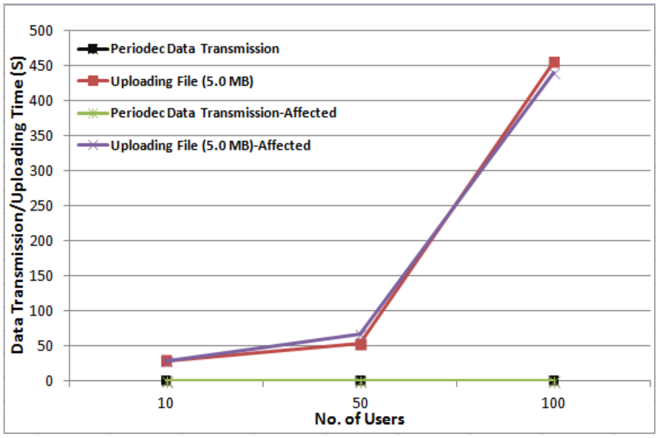


Figure 15. The Effect Uploading Files Scenario on Periodic Data Transmission Scenario

Simulation 6: finally, the mutual effect of combining the traffic of uploading files of size 5.0 MB and the periodic sensors' data transmission according to the signaling flow of Figure 5 is investigated. The USN Gateway connection BW is 1.46 MB/S (HSUBA Category 2). Each type of traffic was simulated separately and then combined together. The results are shown in Figure 15 for separate and combined traffic types. We can conclude that there is almost no mutual effect as the periodic data size is very small and does not require high bit rate. Therefore, the emergency scenarios reliability will not be affected by the simultaneous transmission of other files.

V. Conclusion

Despite that e-Health is one of the promising services in NGN, there is no complete or detailed solution to provide this service. In this paper, we tried to propose a complete solution, including both the architecture and the inter-entities signaling to provide e-Health services in NGN based on the IMS and using the SIP protocol to fulfill the service requirements of the USN applications according to the ITU recommendation. A new architecture is introduced. It uses the existing IMS-based NGN functional entities adding to them new ones such as the e-Health AS integrated with a WS and the EHR. Detailed scenarios are presented showing the complete execution of the service and the interaction between the different entities.

The evaluation of the proposal proves the ability to implement the proposed e-Health scenarios and the reliability of the new signaling model. The results show: first, the initialization time and the emergency detection time are very short (0.64 S, 0.21 S), respectively, even under low BW and high number of active users. Second, the file uploading time is affected essentially by the file size and the available BW. However, since file transfer is not delay sensitive, this is not a major issue if high BW connections are not available. Finally, combining different types of traffic together does not have a significant effect on the performance of the system.

REFERENCES

[1] [ITU-T Y.2221] ITU-T Recommendation Y.2221 (2010), “Requirements for support of ubiquitous sensor network (USN) applications and services in the NGN environment”.

[2] E. J. Ko, H. J. Lee, and J. W. Lee, “Ontology-based context Modeling and reasoning for U-HealthCare,” Ieice Transactions on Information and Systems, vol. E90d, Aug 2007, pp. 1262-1270.

[3] M. Mitchell, C. Meyers, An-I Wang, and G. Tyson, “Context Provider: Context Awareness for Medical Monitoring Applications”, Conference Procedding IEEE Engineering in Medicine and Biology Society, EMBC, doi: 10.1109/IEMBS.2011.6091297, 2011, pp. 5244-5247.

[4] S. Bhattacharyya, R.A. Saravanagru, and A. Thangavelu, “Context Aware Healthcare Application,” IJCA - International -Journal of Computer Applications, vol. 22, no. 3, May 2011, pp. 7-12 .

[5] J. Kim, J. Jeong, S. Nam, and O. Song, “Intelligent Service Enabler based on Context-Aware in Next Generation Networks,” Proceedings of the 2008 International Symposium on Parallel and Distributed Processing with Applications, doi: 10.1109/ISPA.2008.118, Dec 2008, pp. 802-806.

[6] F. Paganelli, E. Spinicci, and D. Giuli, “ERMHAN: A Context-Aware Service Platform to Support Continuous Care Networks for Home-Based Assistance,” International Journal of Telemedicine and Applications, vol. 2008, no. 4, doi:10.1155/2008/867639, Jan 2008, pp. 802-806.

[7] F. Paganelli and D. Giuli, “An Ontology-Based System for Context-Aware and Configurable Services to Support Home-Based Continuous Care,” Ieee Transactions on Information Technology in Biomedicine, vol. 15, Mar 2011, pp. 324-333.

[8] M. C. Domingo, “A Context-Aware Service Architecture for the Integration of Body Sensor Networks and Social Networks through the IP Multimedia Subsystem,” IEEE Communications Magazine, vol. 49, Jan 2011, pp. 102-108.

[9] M. Strohbach, J. Vercher, and M. Bauer, “A Case for IMS, Harvesting the Power of Ubiquitous Sensor Networks,” IEEE Vehicular Technology Magazine, vol. 4, no. 1, doi: 10.1109, Mar 2009, pp. 57-64.

[10] J. Vercher *et al.*, “Ubiquitous Sensor Networks in IMS: an Ambient Intelligence Telco Platform,” in Proc.ICT Mobile Summit, no. 20, Jun 2008.

[11] D. Singh *et al.*, “IP-based Ubiquitous Sensor Network for In-Home Healthcare Monitoring,” International Multimedia, Signal Processing and communication Technologies (IMPACT’09), doi: 10.1109/MSPCT.2009.5164210, Mar 2009, pp. 201–204.

[12] Y. B. Kim, “u-Healthcare Service Based on a USN Middleware Platform and Medical Device Communication Framework,” 5th International Conference on Intelligent Computing (ICIC 2009), vol. 5754, doi: 10.1007/978-3-642-04070-2_76, 2009, pp. 706-714.

[13] A. M. Said and A. W. Ibrahim, "NEW e-health signaling model in the NGN environment," in e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on, doi: doi: 978-1-4577-2038-3/12Oct 2012, pp. 391-394.

[14] ETSI ES 282 004: “Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture; Network Attachment Sub-System (NASS)”, V3.4.1 (2010-03).

[15] ETSI ES 282 003: “Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control Sub-System (RACS): Functional Architecture”, V2.0.0 (2008-05)

[16] [ITU-T Y.2021] ITU-T Recommendation Y.2021 (2006), IMS for Next Generation Network.

[17] Y. Kitatsuji, Y. Noishiki, M. Itou, and H. Yokota, “Service Initiation Procedure with On-demand UE Registration for Scalable IMS Services”, the 5th International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2010), 2010.

On Throughput Performance and its Enhancement in Mobile Ad Hoc Networks (MANETs)

Mouna Abdelmoumen, Mariem Ayedi
and Mounir Frikha

Sup'com
Ariana, Tunisia
Email: {mouna.abdelmoumen, ayedi.mariem, m.frikha}@supcom.rnu.tn

Tijani Chahed

Institut Mines-Telecom; Telecom SudParis
Paris, France
Email: tijani.chahed@telecocm-sudparis.eu

***Abstract*—In this work, we investigate the performance of Mobile Ad Hoc networks (MANETs) in terms of the throughput achieved in transferring packets from source to destination. First, we study the relationship between mobility and routing and define metrics that enable us to derive an analytical expression for the throughput. Secondly, we validate this expression, via simulations, as a function of several mobility patterns, as well as routing protocols, for various nodes speeds. Eventually, we propose the use of additional fixed relays so as to enhance the throughput performance in case of ill-behaved mobility schemes.**

***Keywords*—MANETs; mobility models; routing protocols; throughput model; additional relay proposal.**

I. Introduction

Data traffic transfer in Mobile Ad Hoc Networks (MANETs) requires the existence of a path between source and destination. This path is established based on the use of intermediate nodes which act as relays. As these nodes are mobile, network connectivity can change at any time. And so, the performance of the network, in terms of throughput for instance, is largely dependent on the varying topology of the network, which itself depends on the nodes mobility pattern, as well as the used routing protocol.

Several works studied the impact of mobility on MANET performance. For instance, Grossglauser and Tse [1] showed that the mobility of nodes increases the throughput between source and destination. N. Sadagopan, F. Bai, B. Krishnamachari and A. Helmy [2] defined a connectivity-oriented metric, namely path duration, to analyse the effect of mobility on the connectivity graph between the mobile nodes. They developed a simple first-order model that showed that the throughput is in a strong linear relationship with the reciprocal of the average path duration. As of its relationship to routing, the same study showed how mobility impacts the performance of reactive routing protocols in MANETs.

Despite the fact that these works focused on the relationship between mobility and MANETs performance, they did not give explicit details on the relationship between mobility and routing in this context. In the present work, we focus on this relationship between mobility and routing, and define mobility and routing oriented metrics, namely mobility and routing path durations and path absence durations, which would enable us to model the throughput achieved in transferring packets between source and destination. This model will be next validated through simulations by considering different mobility patterns and routing protocols. Eventually, and in the case of poor throughput performance due to network fragmentation, we propose a new solution based on the deployment of additional fixed relay nodes that would maintain mobility and routing paths for longer and hence enhance the overall network performance.

The remainder of this work is organized as follows. In Section II, we focus on the relationship between mobility and routing and derive an expression for the throughput based on mobility and routing oriented metrics. In Section III, we evaluate these metrics and validate our throughput model by means of comparison between analytical and simulation results. In Section IV, we present our proposal for additional fixed relay nodes deployment and quantify its impact on the network performance. Section V concludes the paper.

II. Model

In order to study the network performance, in terms of throughput, one needs to characterise the paths established between the source and destination. In order to do so, we first focus on the relationship between mobility and routing.

A. Mobility-Routing Relationship

When two nodes i and j come within each other communication range, a so-called *mobility* link, denoted by $L_m(i,j)$, is established. The path between the source and destination is a succession of such links, whose creation/destruction is function of nodes encounters/dis-encounters, and hence mobility.

Once this mobility-based path is established between the source and destination, and before the effective transfer of information between them, the routing protocol exchanges some information, such as routing table update, route request/response, etc, so as to enable the source and destination nodes to learn about the existence of a path between them. A *routing* link between nodes i and j, denoted by $L_r(i,j)$, will

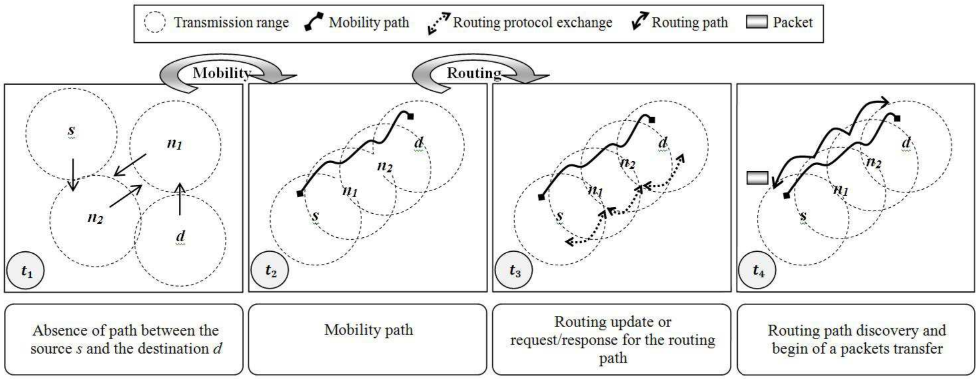

Transmission range
Mobility path
Routing protocol exchange
Routing path
Packet
Mobility
Routing
s
n_1
n_2
d
t_1
t_2
t_3
t_4
Absence of path between the source s and the destination d
Mobility path
Routing update or request/response for the routing path
Routing path discovery and begin of a packets transfer

Fig. 1. Mobility-Routing relationship.

thus be created on top of the mobility link, $L_m(i,j)$. Figure 1 illustrates this situation.

Both mobility and routing paths are composed of time-varying sub-, or unitary, paths relating successive nodes between source s and destination d pairs.

The mobility and routing unitary paths, denoted by up_m and up_r, respectively, between nodes s and d, are composed of $k-1$ consecutive mobility or routing links, respectively, and are defined in a similar fashion as $up_a(s,d,t_c,t_v) = \{L_a(s,n_1),...,L_a(n_{k-2},d)\}$, where index a can be replaced by m for mobility and r for routing. t_c is the establishment or discovery time of the mobility or routing unitary path, respectively, and t_v is the corresponding break or interruption time. Also, we define $\forall q$, $up_{a_q}(s,d) = up_{a_q}(s,d,t_{c_q},t_{v_q})$. Then, the mobility and routing paths, denoted by P_m and P_r, respectively, between nodes s and d, are composed of successive unitary paths: $P_a(s,d,t_c,t_v) = \{up_{a_g}(s,d),...,up_{a_l}(s,d)\}$. $P_a(s,d,t_c,t_v)$ is the mobility or routing path between s and d established at time $t_c = t_{c_g}$ and interrupted at time $t_v = t_{v_l}$ with $\forall q \in [|g+1,l|]; t_{c_q} \simeq t_{v_{q-1}}$ or $t_{c_q} - t_{v_{q-1}} \leq \varepsilon$.

The absence of mobility, as well as routing paths AP_m and AP_r, respectively, between s and d corresponds to the absence of successive mobility and routing links between s and d for a period of time larger than ε. So, we have, $AP_a(s,d,t_i,t_f) = (up_{a_i}(s,d), up_{a_f}(s,d))$ where $t_{c_f} \gg t_{v_i}$.

B. Metrics

To quantify the mobility and routing paths, we define, first, the link duration, $LD_a(i,j,t)$, observed at time t, as the longest time interval, $[t,t']$, during which $L_a(i,j)$ exists. Based on the work of N. Sadagopan et al. [2], we define the mobility and routing path duration, $PD_a(s,d,t_1,t_2)$, which is equal to:

$$\sum_{up_{a_q}(s,d)\in P_a(s,d,t_c,t_v)} \min_{1\leq h\leq k_q} LD_a(n_h,n_{h+1},t_{c_q}) \qquad (1)$$

where $\forall\ up_{a_q}(s,d)$; $n_1 = s$, $n_{k_q} = d$ and k_q is the number of nodes of the unitary path.

The duration of the path absence is simply given by:

$$APD_a(s,d,t_i,t_f) = t_f - t_i \qquad (2)$$

For an observation duration denoted by $T = [t_{begin}, t_{end}]$, we define three sets. The first set, denoted by $P_a(s,d,T)$, contains all of the paths between s and d observed during T; $\{P_{a_z}(s,d) = P_a(s,d,t_{c_z},t_{v_z}); t_{c_z}, t_{v_z} \in T, t_{c_z} \leq t_{c_{z+1}} \forall z \gg 0\}$. The second set, denoted by $AP_a(s,d,T)$, contains all the absences of paths between s and d observed during T; $\{AP_{a_z}(s,d) = AP_a(s,d,t_{i_z},t_{f_z}); t_{i_z}, t_{f_z} \in T, t_{i_z} \leq t_{i_{z+1}} \forall z \gg 0\}$. The third set, denoted by MP_{SD}, contains all source-destination pairs between which a mobility path will be investigated.

For the observation duration T and for the MP_{SD} set, we derive the following average metrics. The average path duration, $\overline{PD_a}$, is equal to :

$$\frac{\sum_{(s,d)\in MP_{SD}} \overline{PD_a(s,d)}}{Card(MP_{SD})} \qquad (3)$$

where $\overline{PD_a(s,d)} = \frac{\sum_{P_{a_z}(s,d)\in P_a(s,d,T)} PD_{a_z}(s,d)}{Card(P_a(s,d,T))}$ having $PD_{a_z}(s,d) = PD_a(s,d,t_{c_z},t_{v_z})$ and where Card is the number of elements in the set.

The average path absence duration, $\overline{APD_a}$, is equal to:

$$\frac{\sum_{(s,d)\in MP_{SD}} \overline{APD_a(s,d)}}{Card(MP_{SD})} \qquad (4)$$

where $\overline{APD_a(s,d)} = \frac{\sum_{AP_{a_z}(s,d)\in AP_a(s,d,T)} APD_{a_z}(s,d)}{Card(AP_a(s,d,T))}$ having $APD_{a_z}(s,d) = APD_a(s,d,t_{i_z},t_{f_z})$.

C. Throughput Model

We assume a full buffer case wherein source s has continuously data to transfer to destination d during observation duration T at transmitting rate (traffic rate) $r_{traffic}$. When the mobility path between s and d is established, the routing path can be set on top of it and hence data transfer can take place. As the routing path constitutes the effective opportunity to exchange data between source and destination nodes, the transfer of data is performed only during the routing path period represented by $\overline{PD_r(s,d)}$ and is interrupted during the path absence duration accounted for by $\overline{APD_r(s,d)}$. Figure 2 illustrates this situation.

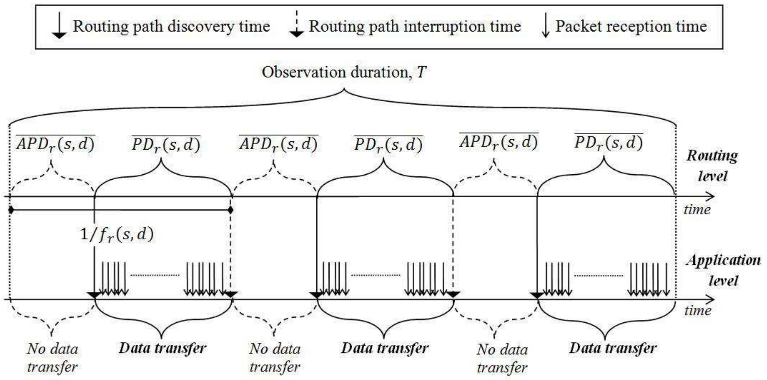


Fig. 2. Relationship between the routing level and the packet transfer.

Hence, for the source-destination pair (s, d), the observation duration T is divided as follows:

$$T = T_{Transfer} + T_{No\,Transfer} \tag{5}$$

where $T_{Transfer}$ and $T_{No\,Transfer}$ are the total durations during which a data transfer takes place and is interrupted, respectively.

Following Figure 2, we have:

$$T_{No\,Transfer} = \overline{APD_r(s,d)} * f_r(s,d) * T \tag{6}$$

where $f_r(s,d)$ is the routing path discovery frequency which is equal to $\frac{1}{\overline{PD_r(s,d)}+\overline{APD_r(s,d)}}$. Hence, (6) becomes:

$$T_{No\,Transfer} = \frac{\overline{APD_r(s,d)}}{\overline{PD_r(s,d)} + \overline{APD_r(s,d)}} * T \tag{7}$$

Based on (5) and (7), we obtain:

$$T_{Transfer} = (1 - \frac{\overline{APD_r(s,d)}}{\overline{PD_r(s,d)} + \overline{APD_r(s,d)}}) * T \tag{8}$$

Now, we suppose that the total quantity of information transfered between nodes s and d during T is $D(s,d)$. The connection throughput, denoted by $Th(s,d)$, is equal to $\frac{D(s,d)}{T}$. And so,

$$Th(s,d) = (1 - \frac{\overline{APD_r(s,d)}}{\overline{PD_r(s,d)} + \overline{APD_r(s,d)}}) * \frac{D(s,d)}{T_{Transfer}} \tag{9}$$

In addition, we assume that the packet reception rate, denoted by $r_{recp}(s,d)$, is different from the packet generation rate, denoted by $r_{gen}(s,d)$. This difference is due to many factors such as the number of links of the routing path, the transmission conditions, the inter-frame waiting times, as well as the routing path repair duration. Hence, we have:

$$\frac{D(s,d)}{T_{Transfer}} = \frac{r_{gen}(s,d)}{r_{recp}(s,d)} * r_{traffic} \tag{10}$$

The connection throughput is thus equal to:

$$(1 - \frac{\overline{APD_r(s,d)}}{\overline{PD_r(s,d)} + \overline{APD_r(s,d)}}) * \frac{r_{gen}(s,d)}{r_{recp}(s,d)} * r_{traffic} \tag{11}$$

III. Model Validation and Performance Evaluation

In this section, we evaluate our metrics and throughput model using simulations.

A. Simulation Settings

In order to evaluate the mobility and routing metrics and to validate our connection, throughput model, we consider the following mobility models, routing protocols and network settings. For the mobility models, we have chosen the following widely-used mobility patterns:

- Random Way Point (RWP), as described by D.B. Johnson and D. A. Maltz [3], is the most widely used mobility model for which the node movement is free of restrictions, both temporal and spatial;
- Smooth Random Mobility Model (SRMM), defined by C. Bettstetter [4], enhances RWP by adding a temporal dependency where speed is changed incrementally in a smooth fashion;
- Graph Based Mobility Model (GBMM) which was presented by J. Tian, J. Hahner, C. Becker, I. Stepanov and K. Rothermel [5] performs as RWP, but it constrains the node movement to a connected graph;
- Manhattan Mobility Model (MMM), evoked by F. Bai, N. Sadagopan and A. Helmy [6], includes all dependencies. It makes use of a map to confine movement to lanes. Moreover, nodes move according to a temporal correlation. The nodes speed is constrained by the speed of the front node in the same lane.

For RWP and SRMM mobility models, the value of the pause time is randomly chosen between $10s$ and $60s$. In addition, we use the maps shown in Figures 3 (a) and (b) for GBMM and MMM, respectively.

In addition, we consider the following widely referenced routing protocols:

- Dynamic Source Routing (DSR) [7] is a reactive routing protocol, where routes are created on demand using two mechanisms: route discovery to find routes and route maintenance to preserve them. It is based on source

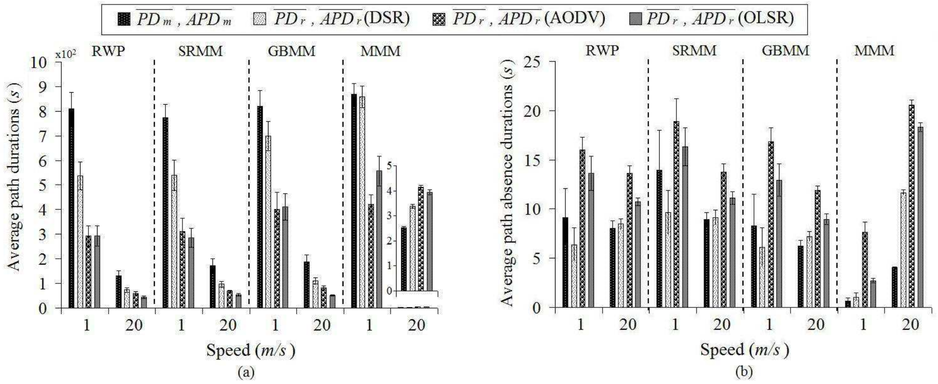


Fig. 4. Average path durations (a) and average path absence durations (b) function of mobility models, nodes speed and routing protocols.

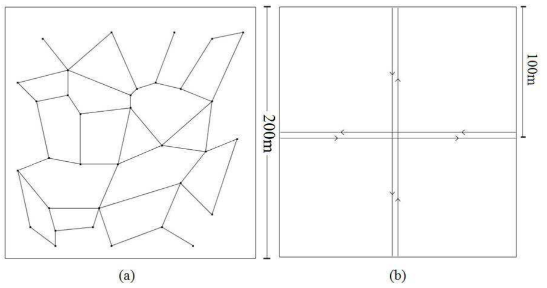


Fig. 3. Maps used for GBMM (a) and MMM (b)

routing whereby all the routing information is maintained and continually updated at mobile nodes;

- Ad-hoc On-demand Distance Vector (AODV) [8] works similarly to DSR using route discovery and maintenance mechanisms. It, however, uses hop by hop routing;
- Optimized Link State Routing (OLSR) [9] is a proactive routing protocol, where the information about the network topology is exchanged by control packets (Hello messages). OLSR makes use of Multi-Point Relays (MPR) nodes to retransmit broadcast messages and hence reduce control packets.

We run simulations over NS-2 (Network Simulator version 2). Simulation duration is taken to be $1000s$. Speeds of 1m/s and 20m/s (equal to 3,6km/h and 72km/h, respectively) are used to mimic the mobility of both pedestrians (low speed) and cars (high speed). Transmission ranges are equal to $100m$.

The traffic rate is 32kbits/s and the data packets size is 96Bytes. Traffic is generated during all simulation duration. So, let the set of the source-destination pairs, MP_{SD}, be $\{(i, i+20)\ \forall i \in [1, N]\}$ where N is the number of nodes.

We use the number of nodes and the simulation area of 10 and 200m×200m, respectively to simulate a high nodes connectivity.

Finally, we generate 20 mobility scenarios for RWP, SRMM, GBMM and MMM based on [10], [11] and [12] tools, respectively.

B. Mobility and Routing Metrics Results

Figure 4 shows the average path durations and the average path absence durations as a function of nodes speed for the different mobility models and routing protocols stated above.

First, we observe that both metrics decrease as speed increases for all mobility models and routing protocols. In effect, when the nodes speed increases, paths are established and broken more frequently and hence the path and absence of path durations decrease. In addition, Figure 4 shows that for high nodes speed, the best values for the metrics are obtained for MMM followed by GBMM, RWP and finally SRMM. When the nodes speed increases, the mobility models performance order changes and becomes GBMM, RWP, SRMM and MMM. The reasons for these results are the following. First, the performance of MMM is due to the map shown in Figure 3(b) where links are only formed if nodes move close to each other in the same or opposite lanes or at intersections which are less probable situation when the nodes speed is high. As a consequence, the network fragmentation occurs which means that some nodes become not reachable by other nodes of the network which leads to large absence of path duration and small path duration. In addition, the performance of random mobility models RWP and SRMM are similar as they allow nodes to move in all directions, and so, links can be formed more frequently than for the MMM model. The good performance of GBMM is due to the fact that it is a mix between restricted and random mobility models. In effect, nodes positions are fixed on the graph as shown in Figure 3(a) and the nodes destination positions are randomly chosen. As a consequence, the probability of link and path establishment is high whatever the network parameters are.

Moreover, we observe that the mobility path durations are larger than the routing ones and, on the contrary, the absence of routing path durations are larger than the mobility ones for all mobility models and nodes speed. The reason is that mobility paths persist more and are less sensitive to interruptions than the routing ones.

Finally, we observe that AODV is the worst routing protocol as it has the lowest values of path durations and largest absence of path durations. The best routing protocol for such

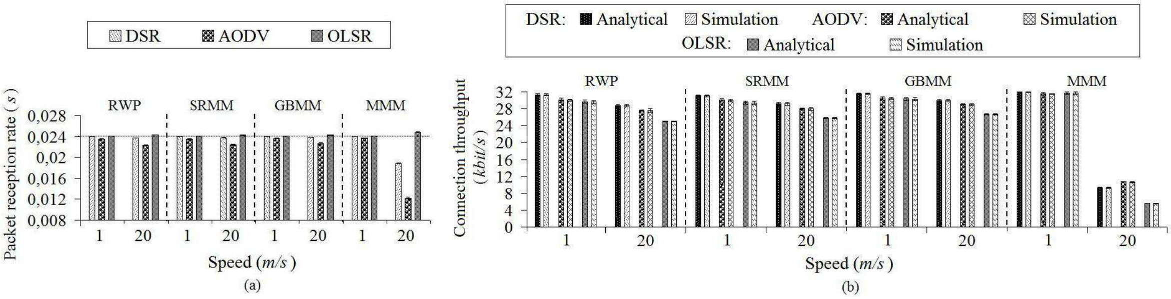


Fig. 5. Packet reception rate (a) and connection throughput (b) function of mobility models, nodes speed and routing protocols.

configuration is DSR. As OLSR is a proactive protocol, at each topology modification and/or periodically, messages are broadcasted to update network information. This fact increases the knowledge about the validity of paths. However, as our network has a high nodes connectivity, less updates are made which allows OLSR to be more efficient. DSR and AODV are reactive, and so, latencies characterize the routes discovery mechanism. Due to the maintenance mechanism however, path interruptions can be quickly detected. In particular, DSR uses MAC notification to detect link failure and AODV uses periodic Hello messages which are broadcasted each $2s$. And hence, DSR failure detection mechanism is more efficient than AODV's. As a result, DSR performs better.

C. Throughput Results

Figure 5 shows the average packet reception rate and connection throughput as a function of nodes speed for the different mobility models and routing protocols obtained from the analytical model (see 11) and from the simulations.

First, we observe that, in almost all cases, packet reception rate is close to packet generation rate. The reason is that as shown for the routing metrics, discovered paths last a longer period of time before breaking. This allows a constant reception rate of packets at the destination. In addition, we observe from Figure 5(b) that the throughput performance follows the mobility and routing metrics as explained above. Moreover, as can be seen from Figure 5(b) throughput under GBMM is the largest among all routing protocols. AODV and DSR work better than OLSR. Those observations are due to the results obtained for mobility and routing metrics discussed above. Furthermore, we observe that throughput reaches 32kbit/s peak and it is low for MMM at high speed. These performances follow, again, the routing metrics: when the routing path duration is high compared to the absence path duration, the network performance is at its best. When the nodes speed decreases, the MMM mobility model works bad as the network fragmentation occurs.

Last, but not least, we observe that our analytical connection throughput model follows closely the values obtained by simulations, as shown in Figure 5(b).

IV. Proposal for Enhancing Throughput Performance

As shown in the previous section, for MMM at high speed, when network fragmentation is frequent, paths cannot be available for a large duration and the network performs poorly. On the contrary, when mobility enables more path establishment opportunities, as in the case of GBMM, throughput achieves a better performance. M. Abdelmoumen, I. Arfaoui, M. Frikha and T. Chahed [13] proposed the use of additional fixed relay nodes so as to improve the network performance by increasing the opportunities of establishing paths and preserving them. We next reproduce its architecture and its impact on enhancing throughput performance for the case of MMM mobility model.

A. Number and Position of Additional Relays

The number of these additional relay nodes must be large enough so as to improve the network performance, but must not exceed a certain limit so as not to overload the network. By trial and error, we fix this number to around 20% of the total number of nodes in the network.

As of their positions, they depend on the (instantaneous) topology of the network. For the special case of MMM model at high speed, as the transmission range of the nodes is sufficiently large compared to the simulation area (100m and 200m×200m, respectively) and with reference to the MMM map (see Figure 3(b)), we choose to fix relay nodes at the positions shown in Figure 6, so as to cover all possible nodes positions and to have a continuous transmission link during the simulation duration.

B. Performance of Proposal

Figures 7 and 8 show the mobility and routing metrics and the average connection throughput, respectively, before and after the use of the additional relay nodes. For comparison, we also show the old values of the studied metrics.

As shown in Figure 7, the mobility path duration increases and the absence of mobility path duration decreases a little with the use of the additional relay nodes. In effect, at high nodes speed, the mobility and routing metrics do not yield a large improvement because nodes move fast enough to have a high connection/disconnection frequency despite of the use of the fixed relay nodes.

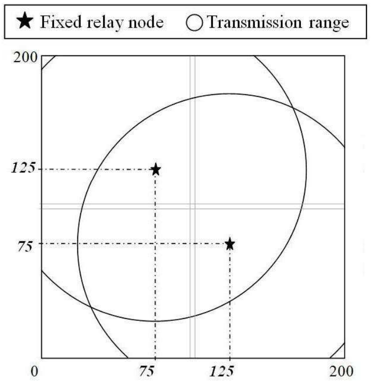


Fig. 6. Relay nodes position for MMM at high speed.

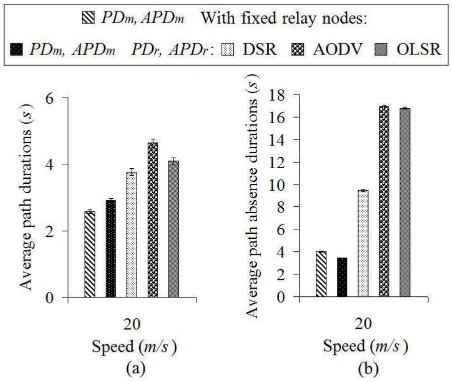


Fig. 7. Average path durations (a) and average of the path absence durations (b) for MMM at high speed without and with fixed relay nodes.

As of throughput, Figure 8 shows that throughput increases with the use of the fixed relay nodes for all mobility models, routing protocols and nodes connectivity. The reason is that packets are lost in smaller number because the absence routing path duration is lower than without the use of the fixed relay nodes. In addition, as explained in the previous section, DSR and AODV work better than OLSR.

V. Conclusion and future work

In this work, we studied the relationship between mobility, routing and MANETs network performance, notably in terms of throughput. We specifically proposed a new model for throughput based on metrics for mobility and routing and validated it in comparison to simulations for various network settings, mobility patterns and routing protocols.

In the case of poor performance, mainly due to network fragmentation, we proposed, and optimized, the use of additional fixed relay nodes which would maintain the overall network connectivity and hence improve the overall throughput performance.

As a future work, we intend to make our study more practical by applying it to a real network.

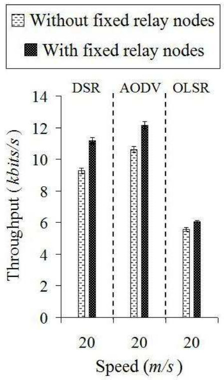


Fig. 8. Network throughput for MMM at high nodes speed without and with fixed relay nodes.

References

[1] M. Grossglauser and D. Tse, ”Mobility increases the capacity of ad hoc wireless networks”, IEEE/ACM Trans. Netw., Aug. 2002, vol. 10, no. 4, pp. 477-486 .

[2] N. Sadagopan, F. Bai, B. Krishnamachari and A. Helmy, ”PATHS: analysis of path duration statistics and their impact on reactive MANET routing protocols”, Proc. ACM MobiHoc, Jun. 2003, pp. 245-256.

[3] D.B. Johnson and D. A. Maltz, ”Dynamic source routing in ad hoc wireless networks”, In Mobile Computing, The Kluwer International Series in Engineering and Computer Science Volume 353, 1996, pp. 153-181

[4] C. Bettstetter, ”Mobility modeling in wireless networks: categorization, smooth movement and border effects”, SIGMOBILE Mob. Comput. Commun. Rev., Jul. 2001, pp. 55-66.

[5] J. Tian, J. Hahner, C. Becker, I. Stepanov and K. Rothermel, ”Graph-based mobility model for mobile ad hoc network simulation”, SS’02, San Diego, Apr. 2002, pp. 337-344.

[6] F. Bai, N. Sadagopan and A. Helmy, ”IMPORTANT: a framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks”, INFOCOM ’03, San Francisco, Mar.-Apr. 2003, pp. 825-835 .

[7] D. Johnson, Y. Hu and D. Maltz, ”The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4”, RFC 4728 Experimental, Feb. 2007.

[8] C. Perkins, E. Belding-Royer and S. Das, ”Ad hoc On-Demand Distance Vector (AODV) Routing”, RFC 3561 Experimental, Jul. 2003.

[9] T. Clauser and P. Jacquet, ”Optimized link State Routing Protocol (OLSR)”, RFC 3626 Experimental, Oct. 2003.

[10] R. Baumann, F. Legendre and P.Sommer, ”Generic mobility simulation framework (GMSF)” , Mobility Model,08 : Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models, Hongkong, China, May. 2008, pp. 49-56.

[11] The CANU Mobility Simulation Environment (CanuMobiSim) [http://canu.informatik.uni-stuttgart.de/mobisim/: Aug. 2014]

[12] The mobility tool generators [http://nile.cise.ufl.edu/important/software.htm: Jun. 2013]

[13] M. Abdelmoumen, I. Arfaoui, M. Frikha and T. Chahed, ”On the performance of MANETs under different mobility patterns and routing protocols and its improvement based on fixed relay nodes”, NTMS, Istanbul, May. 2012, pp. 1-5.

In Vehicle Communication Networks : A Power Line Communication Study and Demonstrator for Infotainment Applications

Fabienne Nouvel, Philippe Tanguy
IETR/INSA, 20 Avenue des Buttes de Coesmes, 35709 Rennes, FRANCE
fabienne.nouvel@insa-rennes.fr,philippe.tanguy@insa-rennes.fr

***Abstract*—The paper deals with in-vehicle communication networks and the use of emerging Power Line Technology (PLC) for infotainment application. It appears that with the increase of Electronic Devices Unit (ECU) both for real time application and infotainment, there is a wire harness bottleneck. So, PLC seems to be a promising in-vehicle network for high data rates applications.After reviewing possible issues, in-vehicle PLC channels and noise measurements are presented. We discuss the Physical Layer (PHY layer) PLC system parameters based on these measurements. Finally, an embedded demonstrator is proposed in order to improve them in a real in-vehicle channel environment.**

Keywords-Intra-vehicle communication; power line communication; impulsive noises; SDR.

I. Introduction

As the Electronic Control Unit (ECU) in vehicles has experienced an exponential demand in the last decades [1][2], the automotive industry has introduced dedicated buses as the Control Area Network (CAN) [5], Local Interconnect Network (LIN) [7],Flexray and Media Oriented Systems Transport(MOST) [7]. CAN and FlexRay have been designed for real-time delivery of messages. However, they do not support high data-rate applications.MOST is used for infotainment applications, and provides a bandwidth of about 50 Mbps. The last version increases the bandwidth up to 150 Mbps and offers a physical layer to implement Ethernet in vehicles. It supports both optical and electrical layers. It allows up to 15 stereo audio channels or MPEG1 channels.Although these networks reduce the amount of wires, they use specific wires and specific protocols. However, the use of the power distribution channels inside vehicles both for power and communication purposes is a promising alternative. It would answer the vehicle requirements namely, cost, decrease of the amount of wires and weight, and offer more flexibility to introduce new applications.

Taking into account both the In-Vehicle Infotainment (IVI) and X-by-wire (X means any mechanical system) requirements, we can observe the necessity to find a limited set of networks which answer to the growing of the multiple applications. These new networks may be able to dynamically optimize their parameters according to the loads on the network (that means active ECUs), the data rate needs (bandwidth sharing) and the state of the vehicle (motor ON/OFF, vehicle in motion, speed), etc. Furthermore, Vehicle to Vehicle (V2V)and Vehicle to Infrastructure (V2X)are currently in active development by automakers like the well-known BMW, Audi, and Volvo [19]. One idea is to propose a new communication protocol which will be compliant both for in vehicle requirements and V2X applications. Among alternatives to existing on-board networks, PLC seems attractive. Many studies are carried out on PLC and focus both on channels and noise in order to optimize the PHY layer and the MAC layer [3].Commercial solutions based on PLC for CAN are provided by Yamar Electronics Ltd. Their PLC-based product families provide maximal data rates of 1.3 Mbps and carrier frequencies in the 1.75 to 13 MHz range. However, they are not yet introduced in vehicles. The current PLC solutions presented by Ferreira et al. [3] show it is possible to achieve data rate more than 50 Mbps, which is consistent with video or audio applications. PLC can be applied also to critical control application. However, the automakers are still reluctant about the transformation of the well known protocols by this PLC solution.Currently, the infotainment area seems to be more open for PLC. In fact, more and more vehicles offer IVI systems and driving technologies, through MOST or wireless connection. The disadvantages of these solutions are the lack of flexibility and the necessary compatibility of devices.One idea is to be able to plug and play any devices offering anTransport Control Protocol/ Internet Protocol (TCP/IP)access, anywhere in the passenger cell of the vehicle. In indoor application, PLC is already used for multimedia application. Latchman et al. [4], the authors provide an overview of the development of the MAC and PHY layers. Repeating functions have been introduced to achieve higher data rate while using frequency bands above 30 MHz.PLC modems may answer theses new challenges, while maintaining bandwidth, data rate and multiplexing flexibility.

In order to introduce PLC in cars, it is necessary to know the channels and the possible scenarios. Section 2 will review previous vehicle networks studies, and focus on PLC. In order to propose PLC solution, it is necessary to study in-vehicle Direct Current (DC) channels to extract the main parameters for the PHY and MAC layers. Section 3 provides a description of the experimentations under in-vehicle DC electrical wires and the main results. Taking into account those in-vehicle measurements, the channel characteristics will allow us to define the best parameters for the signal processing and communication system, like the bandwidth, the modulation, the equalization. The practical approach is discussed in Section 4. In the next step, it seems to be interesting to test several algorithms under real channel while keeping algorithms flexibility. That is why we propose to study a fast prototypingsolution to perform this task using Software Defined Radio(SDR)principle. In this case, the PHY layer is developed in software thanks to Matlab or Labview tools. We used the flexible SDR platform USRP2 [17][18]developed by ETTUS and actually commercialized by National Instrument (NI). This demonstrator is presented in Section 5. Finally, we will conclude the paper in the last section.

II. COMMUNICATION NETWORKS FOR VEHICLES

First of all, the introduction of communication networks was due to that the use of point-to-point communication links was not scalable with the increasing number of electronic components. The demand for efficient networking, including requirements on providing the physical medium (i.e., wires) for communication, has only increased with modern cars being high technology mechatronical systems. This is the main incentive for considering automotive PLC. The Society of Automotive Engineers (SAE) [20] defines four classes (A to D) of automotive communication networks based on transmission speed and applications. The two first classes use low-speed event triggered protocols for low data rate. High-speed real-time communication in Class C and Class D networks relies on high-speed event and time-triggered protocols, respectively.The latter are used for multimedia data, such as audio/video streaming, video games or cameras monitoring, and safety critical applications, such as, e.g., X-by-wire systems, which impose high requirements on availability of resources and reliability of communications.

If we consider these two classes and the reduction of the number of wires, technology originally developed for in-home PLC, namely HomePlug AV, and HD-PLC, seem to be attractive. During the last decades, many studies have been carried out on direct current (12/42 V) voltage for embedded application in vehicles (cars, aircraft) like in [5][6][7]. The results are very promising in that data rates of up to 10 Mbps could be achieved using an approximately 30 MHz bandwidth.Since the medium access control (MAC) protocol is based on the hybrid time-division multiple access (TDMA) and CSMA/CA of HomePlug AV and HD-PLC, it could support both class C and D networks.

Furthermore, Electric Vehicles (EV) may be considered. PLC in EVs has been studiedby Bassi et al. [14].Experiments of Guerrieni et al. [16] using commercial modems from Yamar [16] for PLC in an EV demonstrate reliable communication with data rates of about 1 Mbps.PLC has also been considered for communication between the EV and the charging infrastructure. The IEC 61851 standard defines two charging modes, which require a control pilot signal. Both narrowband and broadband PLC solutions have been proposed for EV to electric vehicle service equipment communication, e.g.,in SAE J2931/2-4 [21]. Recently, the ISO/IEC 15118-3 standard adopts the broadband HomePlug Green PHY as the mandatory PHY/MAC layer technology. The narrowband G3-PLC (ITU-T G.9955) is specified as an optional mode [8].

Among the other solutions for high data rates, we can mention the optical fiber transmission and optical wireless communication.Considering the fiber, this solution has been chosen for the MOST protocol using plastic fiber, from 25 Mbps up to 1 Gbps. The MOST technology is alsoconsidered to be the transmission support for the advanced Driver Assistance Systems (ADAS). One attractive solution is to propose a new modulation scheme like multi-carriers over this channel. A simple single carrier modulation is used but does not exploit all the bandwidth offered by the fiber.On the other hand, Visible Light Communication (VLC) technologycan be used as a medium for data transmission, both for in-vehicle and V2X communication, while achieving high data rates as compared to conventional wireless technologies like for example Wi-Fi or Wimax. In the early 2000s, researches started using visible light from LEDs as the medium for communication. VLC communication can achieve about 800Mbps data rate for short range communications. Researchconductedby Deok-Rae Kim et al. [9] demonstrates the possibility of combining VLC and CAN protocol. Furthermore, we can notice the studies on using Orthogonal Frequency Division Multiplex (OFDM) techniques over these two channels [10]. On restriction of VLC solution is the need to have light and preferred line of sight.We will now discuss about the in-vehicle power lines.

III. CHANNEL MEASUREMENTS

Characterization of PLC channels has been reported according to two major scenarios on four vehicles: front to rear, front to front. These configurations represent possible scenarios for infotainment applications. The measurement setup is presented in detail by Tanguy and Nouvel[7]and is represented in Figure 1. The reference points [A..I] represent both the communication and measurement nodes and a channel between point X to Y point is called XY path. The frequency range goes from DCup to 50 MHz.

Taking into account all the measurements, we have observed an insertion loss of about -15 dB up to -36 dB in the considered band. The maximal attenuation occurs with the longest past GF. The coherence bandwidth $BC_{0.9}$ (90% of the maximum sub-carriers autocorrelation) is given from the autocorrelation of the channel frequency obtained with different paths,as proposed Vallejo-Mora et al. [11].

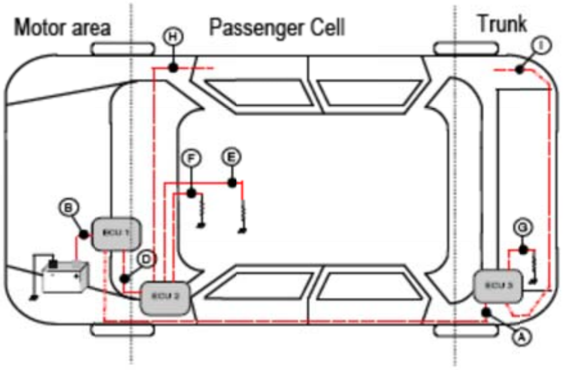


Figure 1.In vehicle test-bed for channel and noise measurements

The channel frequency response H(f) depends of both the sources and loads and is correlated with the transfer function S21 by the approximated relation:

$$H(f) \approx S_{21} \tag{1}$$

As presented in [13], the mean $BC_{0.9}$is greater than 500 KHz and can reach 2 MHz.Furthermore, we have observed its value is not correlated with the path's length. For example, $BC_{0.9}$values are respectively equal to 533 KHz and 4.7 MHz for paths GF (≅ the longest path) on Peugeot 407SW and Renault Laguna vehicles and equal to 2 MHz and 744 KHz for path HD (≅ the shortest path) on the 407 SW and Laguna vehicles.

Additionally, the Root Mean Square(RMS) delay spread is calculated as defined by Ferreira [3]andTlich[13]. In order to define the maximum delay spread a threshold of -30dB has been chosen. If we compute the cumulative density function of the RMS delay, we can observe that 90 % of the channel measurements have a τ_{RMS} delay spread lower than 210 ns with a smallest value of about 50 ns.

If we compare our results with the results obtained in indoor [11], $BC_{0.9}$ in vehicles is as twice as large as those obtained in indoor; the mean $BC_{0.9}$ is of the order of 291.9 KHz in indoor. Keeping the same transmission bandwidth, one can suggest the sub-channel spacing can be reduced. In the case of OFDM technique used, the FFT size may be reduced and the OFDM symbol duration will be shorter. With regard to delay spread, the delay spread obtained in vehicles (maximum value of 242 ns) is two times shorter than the mean value in indoor (0.413 µs). This will allow us to reduce the Cyclic Prefix (CP) length and therefore increase the data rate. These initial results confirm thatthe PLC communication parameters defined for indoor must be optimized for in-vehicle PLC.

IV. PHY PLC PARAMETERS

A. *PLC Transmitted signal*

In our study, Orthogonal Frequency Division Multiplexing (OFDM) [14] technique has been adopted as for indoor PLC, thanks to its high frequency diversity. OFDM modulation can be realized through the IFFT/FFT processing block to which the original stream is applied. Several complementary operations are achieved to the information bits before they are submitted to the IFFT processing. As presented in Section 2, the in-vehicle channels are frequency selective, noisy and multi paths will affect the transmission. For these reasons, OFDM is a good candidate as it divides the bandwidth in flat sub-channels. This solution is applied in other wired or wireless systems and is now well known. The transmitted OFDM waveform can be expressed as:

$$s(t) = \frac{1}{\sqrt{N_p}} \sum_{m=0}^{N_p-1} R_{eal}\left\{c_m \prod(t) e^{2j\pi F_m t}\right\} \tag{2}$$

With NFFT the FFT size, Np= NFFT /2 , c_m the complex symbol on sub-carrier F_m. The sub-carrier spacing Δf. is defined as 1/TOFDM with TOFDM the OFDM symbol duration. During the first experimentation, the OFDM parameters are based on indoor PLC ones like HPAV and HD-PLC [15]. Those standards have been applied in vehicles to measure data throughput on DC channels [16]. The first results demonstrate the feasibility of PLC, and allow about 25 Mbps/s data transmission.

In order to be compliant with vehicle electromagnetic compatibility, we adopt a constant transmit Power Signal Density (PSD) of -60 dBm/Hz in the [4-30] MHz range and 0dBm outside. The additive white gaussian noise has a PSD of -110dBm/Hz. We choose a sub-carrier spacing of 24.4 KHz and a sampling rate of 75 MHz. A bit loading algorithm with a maximum of 10 bits per hertz is accomplished. For the channel, we have considered the channel's measurements, as described in Section 2.

Taking into account those default values, two parameters at the PHY layer are studied and optimized, namely, the guard interval CP and the FFT size.

B. *PLC optimization*

Two parameters have been considered: the CP length and the FFT length, which impact the capacity of transmission given by the equation :

$$R_{OFDM} = \frac{N_{FFT}}{N_{FFT+CP}} \Delta f \left(\sum_{m=0}^{N-1} \log_2 \left(1 + \frac{|H_m|^2 \sigma^2}{\Gamma(\sigma_n^2 + N_{ISI+ICI})} \right) \right) \tag{3}$$

with σthe signal power, $\sigma_n$the noise power, H_m the channel coefficient at frequency m, Γ the margin gain for a BER of 10^{-3}, N the number of used sub-carriers (for example 1148 tones with a FFT size of 3072). The interferences $N_{ISI+ICI}$ are assumed first null. Considering a SNR of 40dB and a BER of 10^{-3}, we have noticed the capacity achieves a maximum of 500 Mbps/s with a CP of 200 ns (15 samples), 186 (14 samples) and 133 ns (10 samples) respectively for paths GF, GH and HD. The higher the channel attenuation is, the higher the value for CP is that maximizes capacity as for high SNR the performances are MOST driven by interference mitigated by long CP.

Then it is possible to combine the CP length adaptation with bit loading algorithm [9], while keeping 1148 used tones. Three different engine states are considered as previously discussed. The theoretical data rate achieved combining the CP optimisation and bit loading is in the range [100 – 200]Mbps. It has been observed the data rate decreases when the engine is turn on, except for the short front paths HD and FD. These results are closed to the measurements carried out by Degardin et al. [6]. These comparisons between measurements and simulations demonstrate the accuracy of our model.

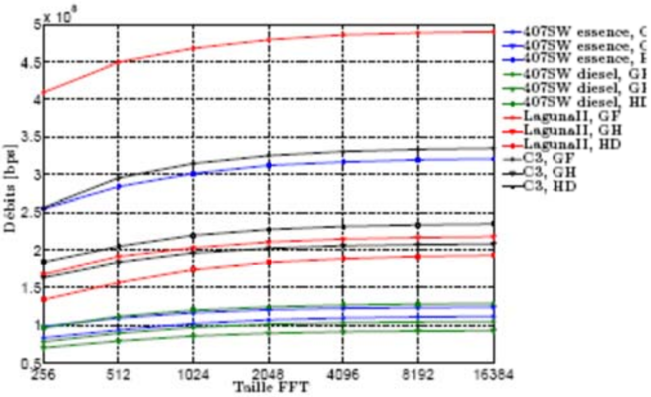


Figure 2.Data rate according FFT size.

The second parameter, the FFT size,has been consideredfor the same scenarios. Taking the $BC_{0.9}$we have obtained, it is obvious the FFT size may be reduced. Four FFT sizes, from 3072/1536 used (Δf = 24,4 KHz<$BC_{0.9}$) down to 128/64 used (Δf = 585 KHz<$BC_{0.9}$) tones are rather interesting. Figure2 gives the data rate we can achieve using those different FFT sizes on different vehicles. The optimal CP is respectively equal to 14 and 50 samples for FFT= 128 and FFT=256, 1024, 3072.

We can observe that when the FFT size increases (FFT > 1024), the ratio ($N_{FFT}/(N_{FFT}$+CP)) is closed to 1, the data rate tends toward a fixed value. One can conclude it is not necessary to increase the FFT size to increase the data rate. If we consider a BER of 10^{-3}, a SNR of 35 dB is necessary for all the FFT sizes while keeping the sub-carrier spacing lower than the $BC_{0.9}$.

Following this system communication analysis, it maybe necessary to integrate these solutions in an embedded network. A demonstrator using the parameters will be presented in next section.

V. VEHICLE POWER LINE SDR PLATFORM

In our vehicular demonstrator, Software Defined Radio approach can be considered as a “wired” communication system, where some of its functional components, such as modulations, equalization, etc., will be implemented in software. This makes it possible to configure the signal according to the requirements of the application and the characteristics of the communication channel (wired or wireless). The software generated signal, thanks to tools like Matlab [21] or Labview [22], combined with the GNU-Radio user interface and according to the parameters defined previously, is then applied to an USRP platform. The platform is detailed in [17]. Next part will review our Vehicle PLC (VPLC) demonstrator using this approach.

A. *VLC demonstrator*

The demonstrator, named VPLC platform, is based on USRP2 boards [19] combined with daughter boards and presented in Figure 3. Using these daughter boards, the possible operation frequency range is very modular (from DC up to 5.9 GHz). The mother board includes ADC, DAC, a FPGA (Xilinx Spartan XC3S2000) and a Gigabit-Ethernet interface with the PC. The two slots of the board are for the front-end daughter boards. The 14-bit ADC and the two 16-bit DAC of the daughter board are independent and can work up to 100 M samples per second. To interface with the DC lines, we use the LFTX and LFRX cards as they allow transmitting and receiving signal from DC up to 30 MHz. These daughter boards include differential amplifiers and low pass filters for antialiasing. The outputs of the platform are linked to the DC lines through a lower band filter and transformer for better isolation.

B. *Results*

The USRP2 TX and RX boards are arranged in the vehicle according to Figure 1 at points G, F, H and D. We will focus on link GF. In Figure4, one can observe the spectrum of the LFTX daughterboard output with a DSP of -80dBm/Hz. In this figure, only the bandwidth [2-12.5] MHz is used; but it is possible to transmit up to 25 MHz. We can observe that there are no notches compared to HomePlug AV standard as we do not have to consider the same electromagnetic constraints. In Figure 5, we observe the received spectrum, through the longest GF path. The signal is affected by the channel and not flat.

Taking into account a referred bit error rate of 10^{-3} and a transmitted power of -80dBm/Hz, we can achieve a mean data rate of about 30 Mbps when the motor is OFF and 7 Mbps when it is ON. This lower result is caused by the lowest SNR when the motor is ON (lower than 5 dB). When we modify the CP, we obtain a little variation according to its length. We respectively obtain 33 Mbps and 30 Mbps with CP=60 and CP=139. If we analyse the best HD path, this last one achieves a mean data rate of 70 Mbps and no errors are detected during the two scenarios. These data rates are compliant with infotainment applications in-vehicle transmission. Furthermore, the PHY layer parameters can be applied for wireless V2X transmission, allowing nearly seamless transmission between in-vehicle and out-vehicle transmission.

C. *Video transmission/ MAC layer*

In order to transmit video or image file, a simple MAC layer has been defined. The stream or file is divided in payloads blocks, called Packet Data Unit (PDU), of 512 up to 2048 bytes. For each PDU, we add a two bytes preamble, a header and a cyclic redundancy code. These PDUs are then inserted in OFDM symbols. One OFDM symbol can include more than one PDU and similary one PDU can be spread over two OFDM symbols. The PHY parameters are summarized in Table I.

With such MAC and PHY layers, it is possible to achieve up to nearly 70 Mbps for the HD path and about 25 Mbps for the longest path GH. The data rate is a bit lower than when we use random bits (Section B). The results show optimisations need to be carried out, as the source and channel coding, the MAC layer. However, it is still large enough for infotainment application.

The results are also limited by the platform as the transmit and received samples are 16 bits length with a maximum 25 M samples /s over the Ethernet link between the laptop and the USRP2 board.

TABLE I. PHY PARAMETERS

Bandwidth	[2-12]MHz
FFT size/used sub-carriers	1024/412
CP length	15 samples
Mapping (for eahc sub-channel)	BPSK to 1024 QAM
Sampling rate	25Msamples/s

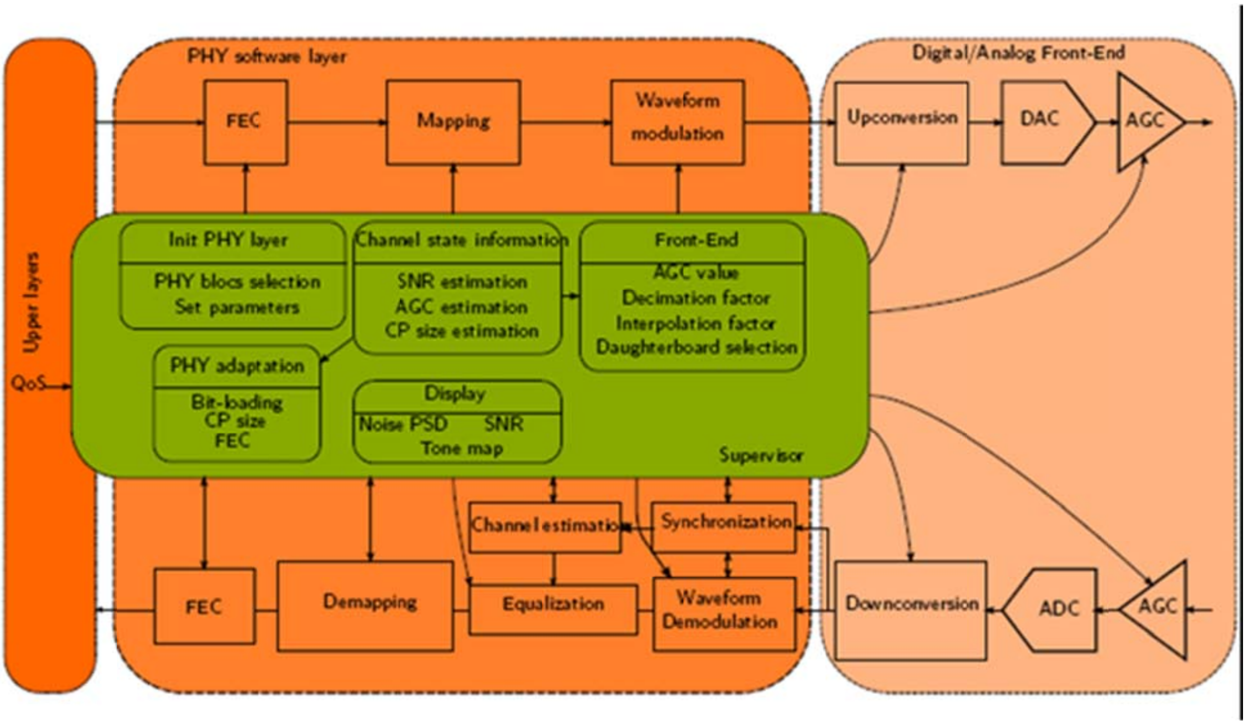


Figure 3.VPLC platform.

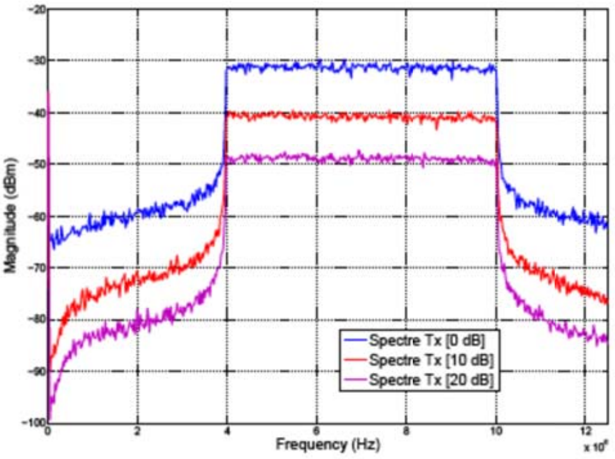


Figure 4.PLC spectrum– TX output, with different attenuations.

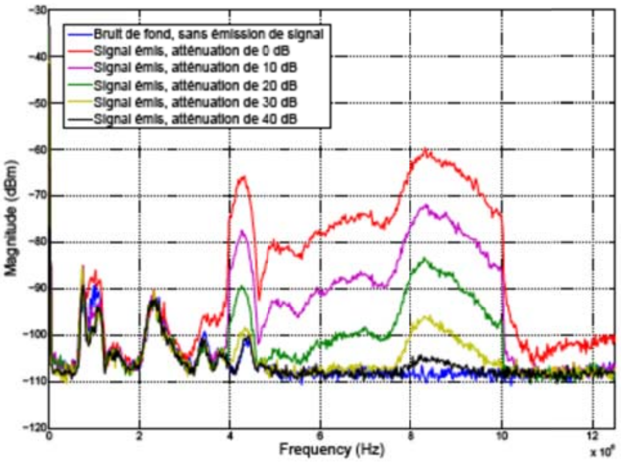


Figure 5.PLC spectrum – RX spectrum, longest path GF, with different attenuation.

VI. CONCLUSION AND FUTURE WORK

In this paper, a PLC communication system has been presented in order to reduce the amount of wiring while offering high data rate for multimedia and infotainment applications.The channel measurements show that OFDM may be applicable, resulting in a good spectral efficiency and high data rate transmission.We have optimized the prefix cyclic length and the FFT sizeof theOFDM signal by taking into account the channel measurements. In order to optimize the network layers, the proposed demonstrator allows us to explore various configurations by adjusting parameters to fully meet expected system requirements. The work in progress focuses on the source coding, the MAC layer and multiple access, using both frequency division multiplexing and time sharing.The objectives are to find the better fit between the PDU and OFDM symbol to achieve a higher data rate.

REFERENCES

[1] G. Len and D. Hefferman, “Vehicles without wires”, Computing & Control Engineering Journal, Volume. N° 12, Iss. 5, pp. 205-221, 2001.

[2] A. J. Van Rensburg and H. C. Ferreira, “Automotive powerline communications: Favourable topology for future automotive electronic trends,” in Proceedings of the International Symposium on Power Line Communications and its Applications (ISPLC), pp. 103–108, 2003.

[3] H. C. Ferreira, L. Lampe, J. Newbury, and T. G. Swart, “Power Line Communications: Theory and Applications for Narrowband and Broadband Communications over Power Lines”, ISBN: 978-0-470-74030-9, 2010.

[4] H. Latchman, S. Katar, L. Yonge, and A.Amarsingh, "High speed multimedia and smart energy PLC applications based on adaptations of HomePlug AV," Power Line Communications and Its Applications (ISPLC), 17th IEEE International Symposium on, pp.143-148, 2013.

[5] W. Gouret, F. Nouvel, and G. El-Zein, “Powerline communication on automotive network,” in Proceedings of the IEEE Vehicular Technology Conference (VTC), 2007, pp. 2545-2549.

[6] V. Degardin, P. Lienard, M. Degauque, E. Simon, and P. Laly,“Impulsive noise characterization of in-vehicle power line,” IEEE Transactions on Electromagnetic Compatibility, vol. 50, no. 4, pp. 861–868, 2008.

[7] P. Tanguy and F. Nouvel, “Power line communication standards for in-vehicule networks,” in Intelligent Transport Systems Telecommunications ITST, pp. 533 –537, 2009.

[8] “Road vehicles—Vehicle to grid Communication Interface—Part 3: Physical and data link layer requirements,” ISO/DIS 15118, 2012.

[9] K. Deok-Rae, Y. Se-Hoon, K. Hyun-Seung, S. Yong-Hwan,and H. Sang-Kook, "Outdoor Visible Light Communication for inter- vehicle communication using Controller Area Network,"Communications and Electronics (ICCE), Fourth International Conference on, pp.31-34, 2012.

[10] L. Peng, “High Data Rate Transmissions over Plastic Optical Fiber:Theoretical Studies and Experiments“, PHD thesis, IETR/INSA Rennes, 2014.

[11] A. B. Vallejo-Mora, J. J. Sanchez-Martinez, F. J. Canete, J. A. Cortes, and L. Diez, “Characterization and evaluation of in-vehicle power line channels” , IEEE Global Telecommunications Conference, pp. 1-5, 2010.

[12] S. D'Alessandro, A.M. Tonello,L. Lampe, "On power allocation in adaptive cyclic prefix OFDM," in Power Line Communications and Its Applications (ISPLC), IEEE International Symposium, pp.183-188, 2010.

[13] M. Tlich, G. Avril, and A. Zeddam, “Coherence bandwidth and its relationship with the RMS delay spread for PLC channels using measurements up to 100 MHz,” , Home Networking, vol. 256, pp. 129–142, 2008.

[14] E. Bassi, F. Benzi, L. Almeida, and T. Nolte, “Powerline communication in electric vehicles,” in Electric Machines and Drives Conference, Miami, FL, USA , pp. 1749–1753,2009.

[15] F. Nouvel and P. Tanguy, "Recent Advance in Power Line Communication for Smart Grid", in “Case study: vehicular networks and architecture ", Chapter 14, CMOS Emerging Technology, 2012

[16] E. Guerrini and G. Dell' Amico, P. Bisaglia and L. Guerrieri, “Bit-loading algorithms and SNR estimate for HomePlug AV”, in IEEE ISPLC 2007, pp. 419-424, 2007.

[17] F. Nouvel and P. Tanguy, “Vehicle Power Line implementation using USRP2 platforms”, in Proc. SDR’2012, Session 2.2, 2012.

[18] Ettus research, llc [Online]. Available: http://www.ettus.com/ [retrieved :07,2014].

[19] Jr. A.Brown, connectivity and the Mobility Industry. SAE International, ISBN of 978-0-7680-4767-7, 2011.

[20] Society of Automotive Engineers [Online]. Available http://www.sae.org/, [retrieved : 06,2014].

[21] Mathworks Company. Available http://www.mathworks.com/index.html,[retrieved :09,2014].

[22] National Instrument Labview tool [Online]. Availablehttp://www.ni.com, [retrieved:09,2014].

Development of a Remote Management System for Automatic Parking Towers through Mobile Devices

Jin-Shyan Lee and Ta-Cheng Chien
Department of Electrical Engineering
National Taipei University of Technology (Taipei Tech.)
Taipei, Taiwan
jslee@mail.ntut.edu.tw and t101318053@ntut.edu.tw

Yuan-Heng Sun
Information & Communications Research Labs
Industrial Technology Research Institute (ITRI)
Hsinchu, Taiwan
gilbertsun@itri.org.tw

Abstract—**Vehicle parking plays a key role in our modern life. Currently, automatic parking towers have attracted much research due to its high space-efficiency. Remote management of a parking tower provides the functions for human to perform monitoring and control via a specific network. From a system point of view, an automatic parking tower is inherently a discrete event system. For such systems, this paper has realized a Java-based management system to provide cross-platform remote access via various devices. In the present approach, Colored Petri Nets (CPNs) are used to model the operated behaviors so as to result in a more compact structure. A prototype of an eight-space parking tower is designed and implemented to show the feasibility of the developed approach. It is believed that the technique presented in this paper could be further applied to large-scale parking systems.**

Keywords—remote management systems; mobile devices; Colored Petri Nets (CPNs); automatic parking towers.

I. INTRODUCTION

Recently, there has been an increasing emphasis on developing consumer electronics, which combines in a synergistic way the classical engineering disciplines of mechanical and electrical engineering and computer science, leading to new kinds of research. In our modern days, car parking is a very crucial issue, and one of the topics in consumer electronics is the investigation of applying the electrical, automation, information, and communication technologies to deal with this problem. In particular, remote management of an automatic parking tower provides the functions for human to perform monitoring and control over a great distance. In real applications, a human operator may use the remote monitoring and control functions to further investigate the parking tower conditions if an alert is launched, or to maintain and repair the parking system if required. However, most of the remote management literature focuses on using a regular PC as the client to access the system. Due to the rapid evolution of mobile phones and tablet PC, the human operator would like to use these modern devices via different web browsers and operation systems to access the server platform at anytime from anywhere [1]-[4]. For such system requirements, this paper has realized a Java-based management system to provide cross-platform remote access via the smart phone, tablet PC, or regular PC.

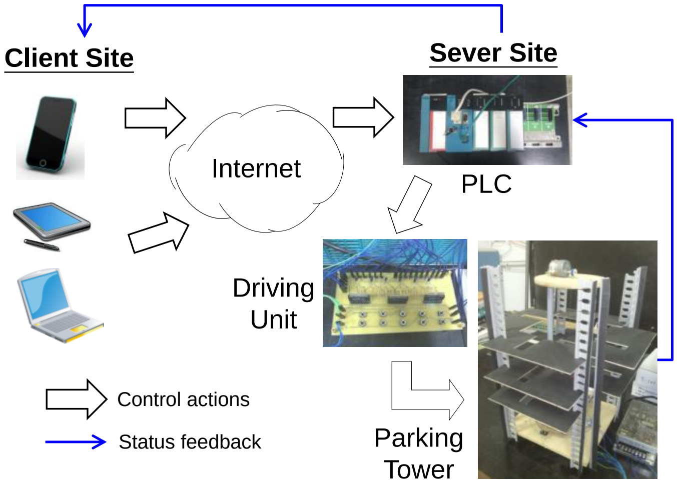


Figure 1. Prototype realization of a parking tower for remote monitoring and control via various devices.

On the other hand, a car parking system is inherently a Discrete Event System (DES), that is, a dynamic system with state changes driven by occurrences of individual events [5]. One way of modeling the DES is using the Petri Nets (PNs) [6]-[8], which have been applied in manufacturing, and more specifically in factory automation, for many years. Their major advantage is the evaluation of all system states before implementation. However, as the number of specifications increases, even simple PN models soon tend to become highly complex. Thus, the use of high-level Petri nets, i.e., Colored PNs (CPNs) has been proposed, leading to a more compact model [9]-[10]. For example, Dotoli and Fanti [11] proposed a Colored Timed Petri Net (CTPN) model to describe in a concise and efficient way the dynamics of an Automated Storage and Retrieval System (AS/RS) serviced by Rail Guided Vehicles (RGVs). Also, Lee and Lee [12] presented a CPN-based modeling and control framework for remotely operated conveyor systems. In their proposed approach, system behaviors were modeling based on the CPN so as to deal with the modeling complexity in large-scale systems with similar behaviors.

Based on the CPN modeling technique, this paper has implemented a prototype of an 8-space parking tower which is automatically controlled by a Programmable Logic Controller (PLC). In our proposed remote management scheme, as shown in Figure 1, the human operator uses the client devices (smart phone, tablet PC, or regular PC) to send control commands to the PLC-based sever through the internet. The commands could be decided according to the status feedback from the server site. Then, the PLC actuates the controlled components in the parking tower via the driving unit. The responses with the status will be fed back to the client site and hence the control loop is closed in this way.

The rest of this paper is organized as follows. Section II briefly introduces the CPN-based modeling scheme. Next, a PLC-based implementation of the remote management system is described in Section III. Then, in Section IV, an example of an eight-space parking tower is illustrated to show the feasibility. Finally, Section V concludes this paper.

II. Modeling via Colored Petri Nets

A. Ordinary Petri Nets

An ordinary PN is identified as a particular kind of bipartite directed graph populated by three types of objects. They are places, transitions, and directed arcs connecting places and transitions. Formally, an ordinary PN is defined as

$$PN = (P,T,I,O,M) \tag{1}$$

where,

$P = \{p_1, p_2, \ldots, p_m\}$ is a finite set of places, where $m > 0$;

$T = \{t_1, t_2, \ldots t_n\}$ is a finite set of transitions, where $n > 0$;

$I : P \times T \to N$ is an input function that defines a set of directed arcs from *P* to *T*, where *N* = {0, 1, 2, …};

$O : T \times P \to N$ is an output function that defines a set of directed arcs from *T* to *P*;

$M : P \to N$ is a marking. An initial marking is denoted by M_0.

A transition *t* is enabled if each input place *p* of *t* contains at least the number of tokens equal to the weight of the directed arc connecting *p* to *t*. When an enabled transition fires, it removes the tokens from its input places and deposits them on its output places. PN models are suitable to represent the systems that exhibit concurrency, conflict, and synchronization. Some important PN properties include boundness (no capacity overflow), liveness (freedom from deadlock), conservativeness (conservation of non-consumable resources), and reversibility (cyclic behavior). The concept of liveness is closely related to the complete absence of deadlocks. Validation methods of these properties include reachability analysis, invariant analysis, reduction method, siphons/traps-based approach, and simulation.

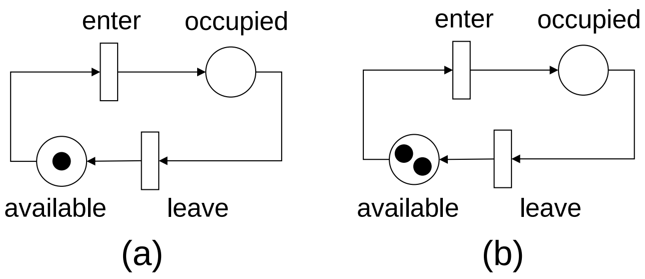


Figure 2. Ordinary Petri net model of the (a) one-space, and (b) two-space parking areas.

Figure 2 (a) shows a simple example of the ordinary PN model for car entering and leaving one parking space in the automatic parking tower. The initial state of parking space is available. Assume a parking tower has *k* parking space, it could be modeled by *k* tokens in the ordinary PN, as shown in Figure 2 (b), where *k*=2. However, using this modeling technique for multiple parking spaces, only the question of “how many” spaces available could be answered, rather than “which” parking space is available. Hence, in order to indicate “which” parking space is available, a more complex PN model, as shown in Figure 3, could be applied. The parking area is organized into two parking spaces, of which the Space-1 is occupied and the Space-2 is available, respectively. Obviously, the model becomes much more complicated as the number of parking spaces increases. Hence, the high-level Petri nets, i.e., CPNs are applied in our work, resulting in a more compact model.

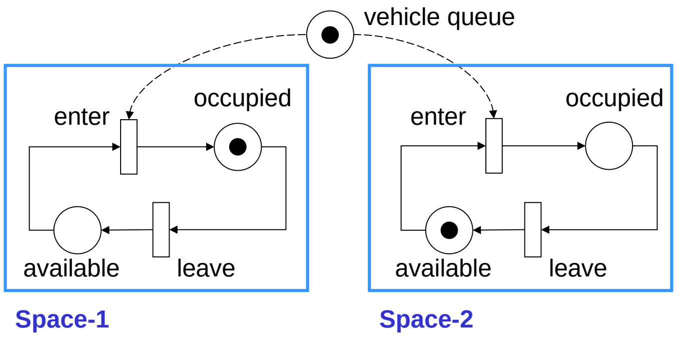


Figure 3. Ordinary Petri net model of a two-space parking area.

B. Colored Petri Nets

A colored PN comprises tokens to which colors are attributed [9]. CPN forms a category of nets whose intuitive perception is less clear than the ordinary PN, but has great value for the modeling of certain complex systems. Formally, a colored PN is defined as

$$CPN = (P,T,C,I,O,M) \tag{2}$$

where,

$P = \{p_1, p_2, \ldots, p_m\}$ is a finite set of places, where $m > 0$;

$T = \{t_1, t_2, \ldots t_n\}$ is a finite set of transitions, where $n > 0$;

C is the color-function; $C(p)$ and $C(t)$ denote the sets of colors associated with place $p \in P$ and transition $t \in T$.

$I(p,t): C(p) \times C(t) \to N$ is an input function that defines a set of directed arcs from *p* to *t*, where *N* = {0, 1, 2, …};

$O(t,p): C(t) \times C(p) \to N$ is an output function that defines a set of directed arcs from *t* to *p*;

$M: C(p) \to N$ is a marking.

Obviously, the CPN is the extension of an ordinary PN. In a CPN, there is a set of colors associated with each place and transition, and a transition can fire with respect to each of its colors. Considering the previous example of the two-space parking tower, the CPN model can be designed as shown in Figure 4 with the elements as follows:

$$P = \{\text{available, occupied}\}$$

$$T = \{\text{enter, leave}\}$$

$$C(\text{available}) = C(\text{occupied}) = \text{SPACE},$$

$$\text{where SPACE} = \{s1, s2\}$$

$$I(p,t) = O(t,p) = ID$$

$$M_0(\text{available}) = s2, M_0(\text{occupied}) = s1$$

Note the *ID* means the identity function, which selects all the items of a basic color domain. Here, the *ID* means the set of SPACE.

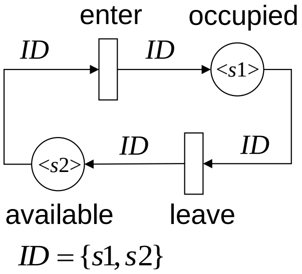


Figure 4. Colored Petri net model of a two-space parking area.

Obviously, the net structure of the CPN (Figure 4) is greatly simplified as compared with the previous ordinary PN model (Figure 3).

III. PLC-Based Realization

A. *Client-Server Architecture*

Figure 5 shows the client-server architecture for implementing the remote management system. On the client side, the remote manager uses a Java-capable web browser, such as Internet Explorer or Firefox, to connect to the web server through the internet. On the server side, an industrial PLC with a built-in Java-capable web server assigned to handle the client requests is employed. Within the PLC, a Java servlet handles user authentication, a Java applet provides a graphical Human-Machine Interface (HMI), and a Ladder Logic Diagram (LLD) performs the detailed operations of the requested tasks. Our choice of using LLD to implement the local operations due to its wide use in industry, while using Java to implement the remote functions because of its object-orientation, portability, safety, and built-in support for networking and concurrency.

B. *Interactive Modeling*

A sequence diagram of the Unified Modeling Language (UML) [13] is applied to model the client-server interaction in the remote management system. As shown in Figure 6, at the first stage, the *Remote Client* sends a HyperText Transfer Protocol (HTTP) request to the *Web Server*. Next, the *Web Server* replies an HTTP response with an authentication web page, on which the *Remote Client* can login to the system by sending a request with user/password. The *Web Server* then invokes a Java servlet to authenticate the user. If the authentication fails, the Java servlet will respond with the authentication web page again. On the other hand, if the authentication succeeds, the Java servlet's response will be a control web page with a Java applet. The Java applet first builds a graphical HMI and constructs a socket on the specified port to maintain continuous communication with the server. Then, the Java applet acquires the system status through the constructed socket and displays it on the control web page iteratively by invoking the *Device Handler* to fetch the sensor states of *Device* objects. After that, the *Remote Client* can issue an action command from the control page to actuate the remote system through the constructed socket. The responses with the status will be continuously fed back to the *Remote Client* and thus the control loop is closed.

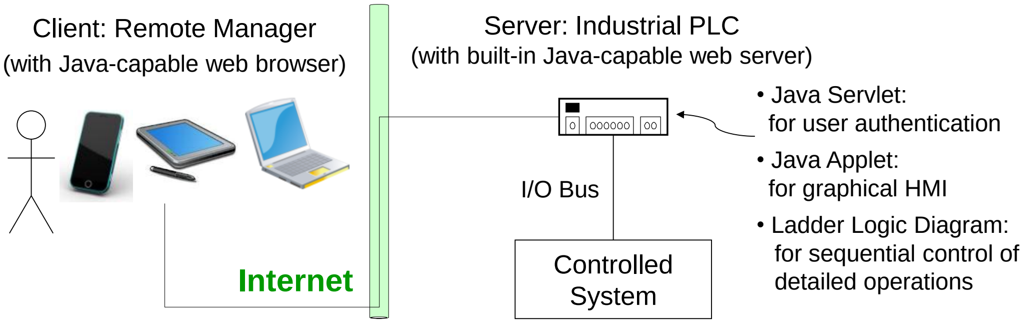


Figure 5. Implementation architecture of the remote management system.

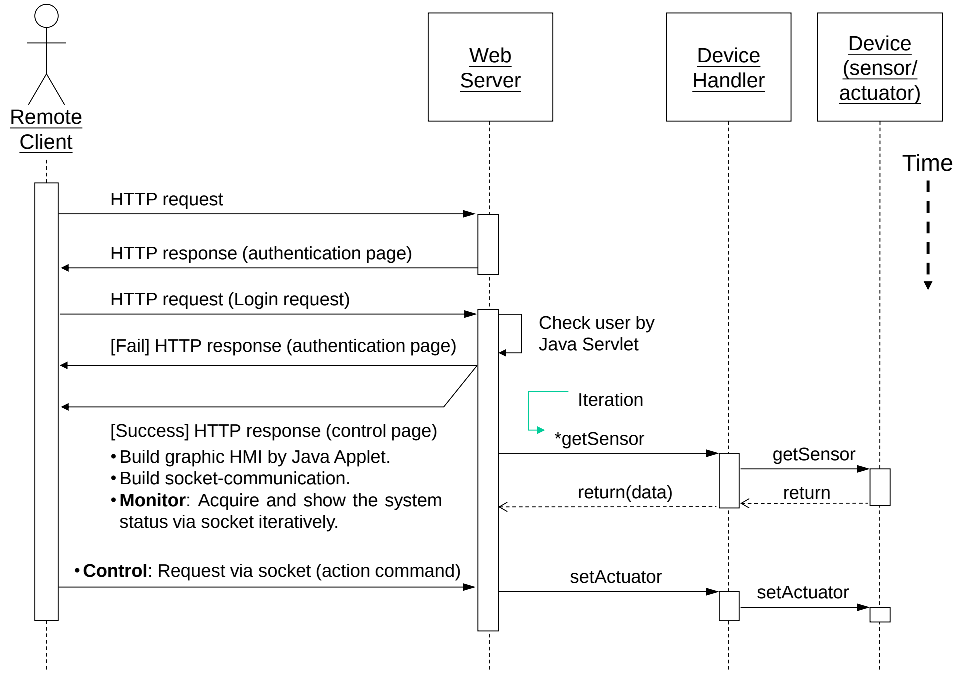

Remote Client
Web Server
Device Handler
Device (sensor/ actuator)
Time
HTTP request
HTTP response (authentication page)
HTTP request (Login request)
Check user by Java Servlet
[Fail] HTTP response (authentication page)
Iteration
*getSensor
[Success] HTTP response (control page)
• Build graphic HMI by Java Applet.
• Build socket-communication.
• Monitor: Acquire and show the system status via socket iteratively.
getSensor
return(data)
return
• Control: Request via socket (action command)
setActuator
setActuator

Figure 6. Interactive modeling with sequence diagram for the remote management system.

IV. An Application Prototype

A. System Description

In a parking tower, the area is organized into a lot of parking spaces. For simplicity, the prototype of an eight-space parking tower is designed and implemented as shown in Fig 7. In the developed tower system, three motors (corresponding to three degrees of freedom) and four limited switch sensors are employed to place the vehicle on a desired parking space. Also, ten control buttons on the driving unit are used to provide local control functions.

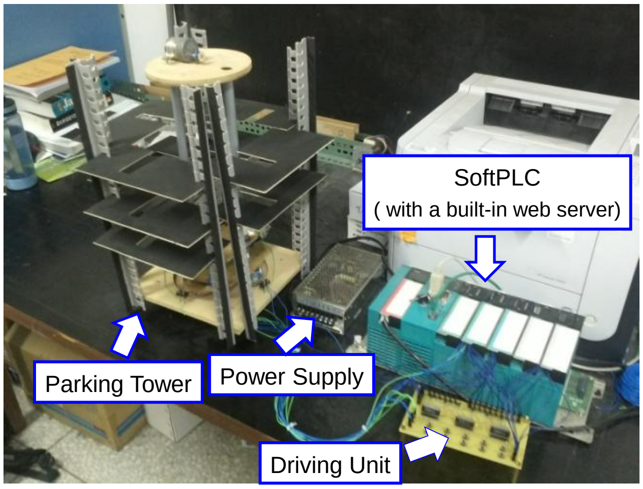


Figure 7. The hardware setup during prototype development.

B. Modeling and Implementation for Local Control

For the operator-issued commands, the CPN model of the eight-space parking system is constructed as shown in Figure 8. The model consists of 4 places, 4 transitions, and 8 color elements (corresponding to eight parking spaces), respectively. Corresponding notation of the PN model is also described in the figure. Also, the related LLD is designed for local control, as shown in Figure 9.

C. Java-Based Realization for Remote Management

To implement the remote monitoring and control functions, we use Java due to its object-orientation, portability, safety, and built-in support for networking and concurrency. The developed server program is located on an advanced PLC (80486-100 CPU) with built-in web server and Java virtual machine so that it can process both HTTP requests and Java programs.

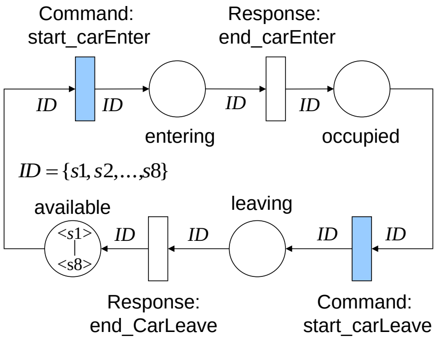
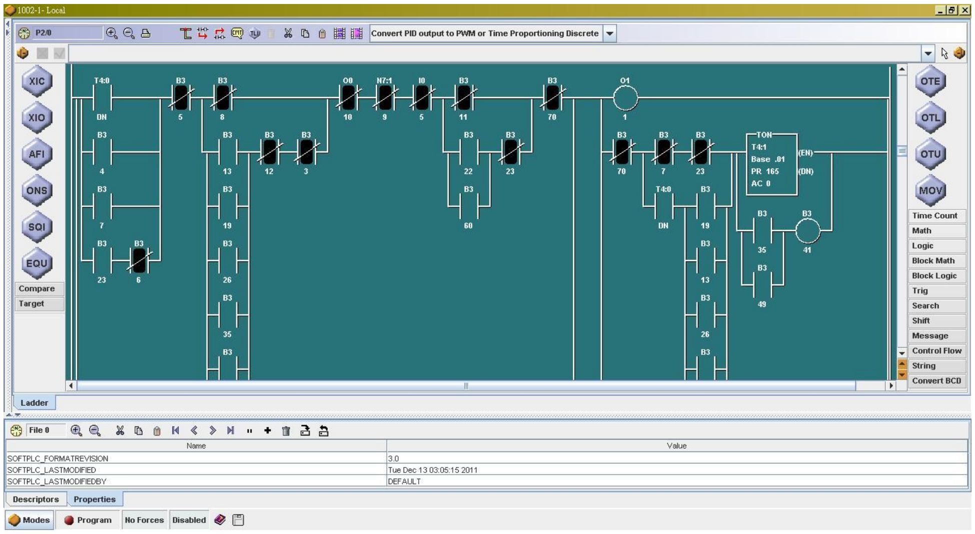
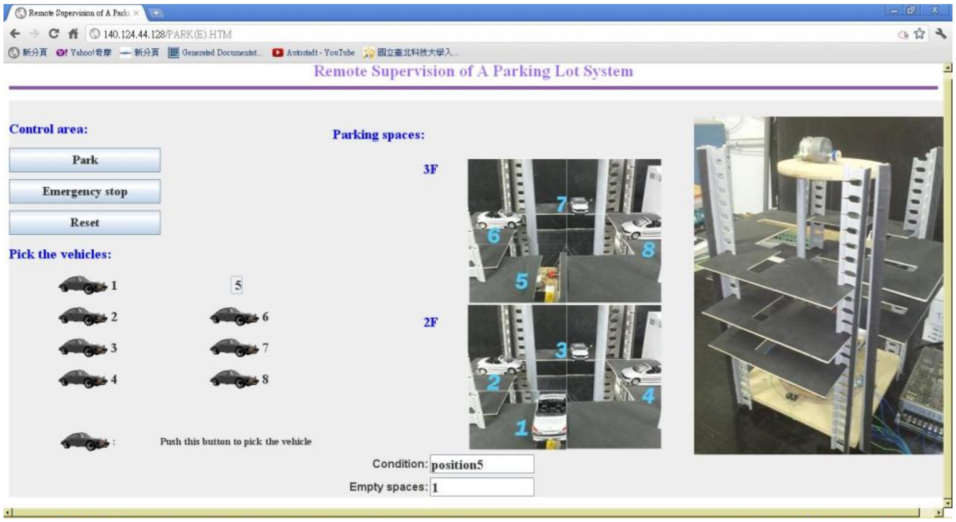


Figure 8. Colored PN model of the eight-space parking system.



Name	Value
SOFTPLC_FORMATREVISION	3.0
SOFTPLC_LASTMODIFIED	Tue Dec 13 03:05:15 2011
SOFTPLC_LASTMODIFIEDBY	DEFAULT

Figure 9. Snapshot of ladder logic diagrams during the PLC implementation.


Remote Supervision of A Parki
140.124.44.128/PARK(E).HTM
Remote Supervision of A Parking Lot System
Control area:
Park
Emergency stop
Reset
Pick the vehicles:
1
2
3
4
5
6
7
8
Push this button to pick the vehicle
Parking spaces:
3F
2F

Condition: position5

Empty spaces: 1

Figure 10. Interactive web page for the remote management of the parking tower.

The developed HMI, shown in Figure 10, is carefully designed to make its web pages more user-friendly. The button control area is placed on the left, and the current status and system message is on the right. The human operator can push the buttons to park a car into a parking space or to take a specific car from the tower.

Figure 11. Remote management of the parking tower via a smart phone.

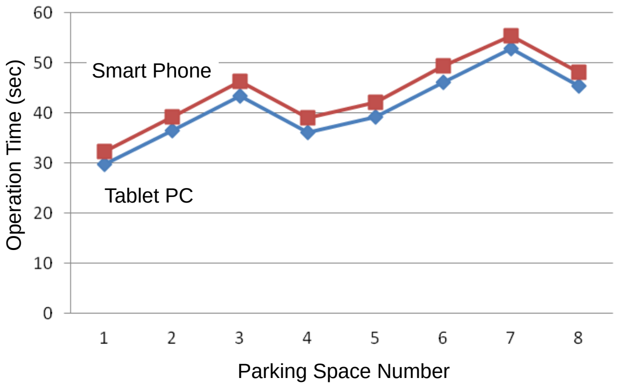


Figure 12. Comparison of using a smart phone and a tablet PC to remotely perform parking (or taking) a car.

Also, the operator can use a smart phone to manage the parking system, as shown in Figure 11. However, due to limitations of computing power and communication bandwidth on smart phones, the consumed time of using the phone to park (or take) a car is a little longer than the time of using a tablet PC. Figure 12 shows the comparison of realistic operation time of parking (or taking) a car between using a smart phone and a tablet PC. Moreover, the operation time of parking (or taking) a car from parking space 7 is the longest, since the space 7 is on the innermost part of the circle as shown in Figure 10. On the other hand, using the space 1 will take the shortest time since it is near the entry.

V. Conclusion and Future Work

This paper was motivated by the requirement of remotely access for automatic parking systems. For such systems, this paper has realized a Java-based management system to provide cross-platform remote access. Moreover, in order to cope with the complexity and realization issues in large-scale parking systems, a systematic development approach is proposed in this paper. In the present approach, the system behaviors are modeling based on colored Petri nets, and then remote management functions are implemented via a Java-based PLC. To demonstrate the practicability of the proposed approach, an application to the eight-space parking tower is illustrated with realization. It is believed that the presented technique could be further extended to large-scale parking towers. Future work will attempt to improve the access-control policy for multiple operators. Also, the issue on how to optimize given problems of parking systems (e.g., how to route cars to the shortest fitting free parking slot) would be addressed in the future.

Acknowledgment

This paper was supported by the Ministry Of Science and Technology (MOST) of Taiwan under the grant 102-2511-S-027-004, and in part by the Ministry Of Economic Affairs (MOEA), Taiwan. Moreover, the authors would like to thank Y-J. Zhuang and Z-N. Zhang, and L-Y. Chen from the Department of Electrical Engineering, National Taipei University of Technology (Taipei Tech.), for their assistance in the prototype implementation.

References

[1] H. S. Ahn, I. K. Sa, and J. Y. Choi, “PDA-based mobile robot system with remote monitoring for home environment,” IEEE Trans. Consumer Electron., vol. 55, no. 3, pp. 1487-1495, Aug. 2009.

[2] G. Paravati, C. Celozzi, A. Sanna, and F. Lamberti, “A feedback-based control technique for interactive live streaming systems to mobile devices,” IEEE Trans. Consumer Electron., vol. 56, no. 1, pp. 190-197, Feb. 2010.

[3] J. S. Lee, “A Petri net design of command filters for semiautonomous mobile sensor networks,” IEEE Trans. Ind. Electron., vol. 55, no. 4, pp. 1835-1841, April 2008.

[4] J. S. Lee and P. L. Hsu, “Implementation of a remote hierarchical supervision system using Petri nets and agent technology,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 37, no. 1, pp. 77-85, Jan. 2007.

[5] B. Hruz and M. C. Zhou. Modeling and Control of Discrete Event Dynamic Systems. Springer: London, 2007.

[6] J. S. Lee, M. C. Zhou, and P. L. Hsu, “An application of Petri nets to supervisory control for human-computer interactive systems,” IEEE Trans. Ind. Electron., vol. 52, no. 5, pp. 1220-1226, Oct. 2005.

[7] H. Zheng, Y. G. Niu, and G. Ciardo, “Modelling and analysis of UPnP AV media player system based on Petri nets,” Int. J. Syst. Sci., vol. 42, no. 9, pp. 1573-1580, Sept. 2011.

[8] J. S. Lee and P. L. Hsu, “Design and implementation of the SNMP agents for remote monitoring and control via UML and Petri nets,” IEEE Trans. Contr. Syst. Technol., vol. 12, no. 2, pp. 293-302, March 2004.

[9] K. Jensen, “Coloured Petri nets,” Lecture Notes Compu. Sci., vol. 254, pp. 248-299, 1987.

[10] K. Feldmann, A. W. Colombo, C. Schnur, and T. Stockel, “Specification, design, and implementation of logic controllers based on colored Petri net models and the standard IEC 1131, Part I: Specification and design,” IEEE Trans. Contr. Syst. Technol., vol. 7, no. 6, pp. 657-665, Nov. 1999.

[11] M. Dotoli and M. P. Fanti, “A coloured Petri net model for automated storage and retrieval systems serviced by rail-guided vehicles: a control perspective,” Int. J. Computer Integrated Manufacturing, vol. 18, no. 2-3, pp. 122-136, March-May, 2005.

[12] J. S. Lee and Y. F. Lee, “Behavior modeling and remote control of industrial conveyor systems via internet,” Proc. IEEE Conf. Industrial Electronics and Applications (ICIEA), Melbourne, Australia, June 2013, pp. 387-392.

[13] G. Booch, J. Rumbaugh, and I. Jacobson. Unified Modeling Language User Guide. Addison-Wesley, 2005.

An Assessment of the Contemporary Threat Posed by Network Worm Malware

Luc Tidy, Khurram Shahzad, Muhammad Aminu Ahmad and Steve Woodhead

Internet Security Research Laboratory
Faculty of Engineering and Science
University of Greenwich
Email: {l.j.tidy, k.shahzad, m.ahmad, s.r.woodhead}@greenwich.ac.uk

***Abstract*—The cost of a zero-day network worm outbreak has been estimated to be up to US$2.6 billion. Additionally zero-day network worm outbreaks have been observed that spread at a significant pace across the global Internet, with an observed infection level of more than 90 percent of vulnerable hosts within 10 minutes. The threat posed by such fast-spreading malware is therefore significant, particularly given the fact that network operator / administrator intervention is not likely to take effect within the typical epidemiological timescale of such infections. This paper presents a classification of wormable vulnerabilities, demonstrating a method to determine if a vulnerability is wormable, and presents a survey into the cause of the reduction of worm outbreaks in recent years, as well as their viability in the future. It then goes on to explore recent wormable vulnerabilities, and points out the issues with operating system security in relation to techniques used by zero-day worms.**

Keywords*—*Cyber Defence; Malware; Network Worm; Zero-Day Worm; Simulation; Modelling

I. INTRODUCTION

As a type of malware that exploits vulnerabilities that have not been patched or acknowledged at the point of an outbreak, with an automatic propagation method that can spread pervasively throughout a network, zero-day worms are particularly virulent. The effects are exacerbated by either a lack of detection or a high speed of propagation [1]. The threat presented by such malware to the Internet, national security and defence systems is therefore significant.

In the first few years of the twenty-first century, there were a number of notable zero-day worm outbreaks [2][3][4] however, since these events the number of zero-day worm outbreaks has reduced. Understanding this reduction, and assessing whether such worm outbreaks are still viable in a modern setting are essential. This paper presents a discussion of historical worm events to ascertain why they occurred, and then discusses the motivations for malware attacks to assess why worm outbreaks have seen this reduction. The paper then presents a discussion on recent wormable vulnerabilities and operating system security, in order to assess whether zero-day worm outbreaks are still viable on the modern Internet.

The remainder of this paper is presented as follows: Section II presents a lexicon as a definition of terms. Section III presents related work, focusing on similar studies into the assessment of potential threats. Section IV presents a discussion on the motivations for carrying out a malware attack. Sections V and VI present a summary of recent wormable vulnerabilities, and addressing the particular issue of operating system security. Finally, the paper is concluded in Section VII.

II. LEXICON

A lexicon has been presented for the clarification of the following terms, owing to their specific use in this paper.

Zero-Day Worm: In this paper, this is defined as a type of malicious software that propagates automatically without human interaction, using a vulnerability that has not been patched or widely acknowledged at the point of an outbreak. In particular, this paper reports findings on fast, random-scanning worms [5]. This is a similar definition to the taxonomy described by Weaver et al. [6], and other published literature (see [2][3][4]).

Wormable Vulnerability: A vulnerability that has the potential for use in worm propagation, as defined by being network accessible, allowing the execution of arbitrary code and whether a not a vulnerability can be exploited remotely. This is in accordance with the model reported by Nazario et al. [7].

III. RELATED WORK

Research into worms and their outbreaks has been reported in three key areas: the classification of worms and wormable vulnerabilities, potential worm outbreak scenarios and the investigation of previous worm outbreaks. In addition, this paper also considers contemporary malware threats.

A. Classification of Worms and Wormable Vulnerabilities

The taxonomy reported by Weaver et al. [6] presents an overall method of classifying worms. The classification is made under the following categories: target discovery, propagation method, activation, payloads, motivations and attackers. Similar categories are reported by Li et al. [8], which classified worms under a number of schemes: target finding, propagation, transmission and payload. Smith et al. [9] also uses the taxonomy reported by Weaver, however, expands this further to consider evasion and detection methods, which incorporate different propagation methods and payloads. For the purposes of this paper, we choose to focus on self-carried worms, or worms that do not require other network traffic in order to propagate.

Another factor of classifying worms is the vulnerability they exploit in order to propagate. As reported by Nazario et al. [7], a wormable vulnerability can be summarised in (1), where wormability, W, is a product of the exploit characteristics, E, population characteristics, P, and the time since the disclosure of the vulnerability to account for development of the worm. Nazario et al. also defines the characteristics of a wormable exploit, as shown in (2), where the exploit characteristics, E, are defined by the fractional population of exploit architecture, f_{E_p}, the fractional availability of an exploit for a given vulnerability, f_{E_a}, the number of chances available to attempt an exploit, E_c, the fraction of exploit reliability, f_{E_r}, the Boolean value of whether the exploit is able to be made remotely, R, if the impact of the vulnerability is execution of code, I_e and if the impact of the vulnerability permits network access, I_n.

$$W = E * P * L \tag{1}$$

$$E = f_{E_p}(f_{E_a} + 0.067)(\frac{E_c - 1}{E_c} + f_{E_r})RI_eI_n \tag{2}$$

Using the key factors reported by Nazario et al., and those reported by Weaver at al., Li et al., and Smith et al., a wormable vulnerability can be summarised in the Boolean equation (3), where a wormable vulnerability, V_w, is determined by not requiring human interaction, H, is network reachable, N_r, provides remote code execution, R, and provides network access, N_a once exploited.

$$V_w = \bar{H} \bullet N_r \bullet R \bullet N_a \tag{3}$$

In addition to the reported work that provides a classification, there are also a number of resources that focus on providing details for known vulnerabilities. One such source is the Common Vulnerabilities and Exposures (CVE) system [10], which provide details for a range of vulnerabilities. The CVE system notes the access vector, for instance if the vulnerability is network reachable or requires human interaction, and the impact if the vulnerability were to be exploited, for instance providing remote code execution or network access. These details provide information in order to assess whether a vulnerability is wormable.

B. Potential Worm Outbreak Scenarios

Potential worm outbreak scenarios often focus on new technologies or methods that a worm may use in order to spread faster. As far as the authors are aware, the first notable instance of this was the work reported by Weaver in 2001 [11], which described a Warhol worm - where using a combination of a list of known vulnerable hosts, known as a hitlist, and by dividing up how each worm scans for new susceptible hosts, known as permutation scanning, the worm increases in virulence. Such methods were seen in the Witty outbreak of 2003 [4], and the second version of Code Red, Code Red II, in 2001, respectively.

Work reported by Staniford et al. [12], presents results on the impact of very fast, what is termed as Flash worms, on a contemporary Internet as of 2004. Using simulation, Staniford estimates that an optimised Flash worm could spread within seconds. Similar fast outbreaks are further corroborated in work reported by Tidy et al. [13], as well as reporting work on other potential scenarios in [5], where a worm uses an intentionally slow phase before switching to a fast, random-scanning method in order to increase the number of infected hosts prior to its fast phase; resulting in an impact similar to having a hitlist.

The work in potential worm outbreaks assume that a wormable vulnerability exists, however, there is limited work in investigating contemporary vulnerabilities in order to determine if they are wormable, and the possible worm outbreaks that could occur.

C. Previous Worm Outbreaks

There have been a number of large-scale zero-day worm outbreaks, most notable of which are the Morris Worm outbreak of 1988 [14], the Code Red outbreak of 2001 [2], the Slammer outbreak of 2003 [3] and Witty outbreak of 2004 [4]. Table I summarises these worms, detailing the platform/service that had the wormable vulnerability, the port/s used for propagation, and the exploit method. This shows that these notable events all used a buffer overflow in order to infect susceptible hosts, propagated using different ports and exploited vulnerabilities on a number of different platforms.

Another reported characteristic of these worm outbreaks centre around their payload. Both the Morris and Slammer worms contained no destructive or directly malicious content as part of its payload. Similarly, the Code Red worm only began to undertake a denial of service attack after it had completed a propagation phase. As reported by Shannon and Moore [4], the Witty worm was the first to carry a destructive payload, overwriting randomly chosen sections of the infected hosts hard drive with the phrase “`(^.^) insert witty message here (^.^)`”.

Owing to the lack of malicious payload in the Slammer worm, the intentional pause in propagation in the Code Red worm and as the Morris worm was described by its author to be designed to gauge the size of the ARPANET, it can be argued that the motivation to release these worms was one of discovery. Similarly, as the Witty worm was the first of these

TABLE I
SUMMARY OF NOTABLE WORM CHARACTERISTICS

Name	Vulnerable Platform/Service	Port/s	Exploit Method
Morris	DECX Sun 3, sendmail finger	25,79	Buffer overflow
Code Red	Microsoft IIS web service	80	Buffer overflow
Slammer	Microsoft SQL Server 2000	1434	Buffer overflow
Witty	Internet Security Systems firewall	Random	Buffer overflow

to carry a destructive payload, it could have been released to assess whether a destructive payload was feasible.

D. Contemporary Malware Threats

Since the large outbreaks at the beginning of the 21st century, the number of large-scale worm outbreaks has decreased significantly. Panda Security [15] reports that worms only constituted approximately 6% of all malware infections in the first quarter of 2013, it is also reported that trojans constitute the majority of the malware infections with 80% of all malware infections being of this type. One of the largest of these is the Zeus trojan [16], which is designed in order to commit fraud by gaining access to banking details on infected hosts and sending these details to the attacker. This is defined by Wilson [17] as cybercrime, or criminal activity that is “enabled by, or that targets computers”.

A return to worm-like characteristics can be seen in the Stuxnet [18] outbreak, which targeted industrial control systems in order to cause damage. It is suggested that Stuxnet is an example of cyberwarfare [19], where the intent was to cause damage to the targeted industrial systems. This is a distinct difference in the cybercriminal activity, as instead of criminal gain the motivation of released Stuxnet was one of causing damage.

IV. Motivations for Malware Attacks

One of the main factors in understanding malware outbreaks is the motivation of the attacker. A difference in motivation can influence the choice of malware that an attacker will choose, given that different malware is more effective at certain tasks than others. In the case of worm outbreaks, this is demonstrated by the reduction in events, owing to a change in the motivation of attackers. Figure 1 illustrates this change, plotting the trend of worm prevalence against time, along with three categories of attacker motivation: experimentation or discovery, cybercrime and cyberwarfare.

Up to the first few years of the 21st century, the use of malware was comparatively in its infancy, and the notable worm outbreaks during this period can be argued to have been for experimental purposes, with the main motivation of the attacker to see if they are feasible; or in the case of the Morris worm to measure the size of the ARPANET. From around 2004 onward, the use of malware for cybercrime has increased. Such criminal activity, as shown by the prevalence of trojans like Zeus [16], has focused on gaining access to confidential data or disrupting services, such as a Distributed Denial-of-Service attack (DDoS) [20], through the use of controlling a large number of machines through a botnet created using a trojan.

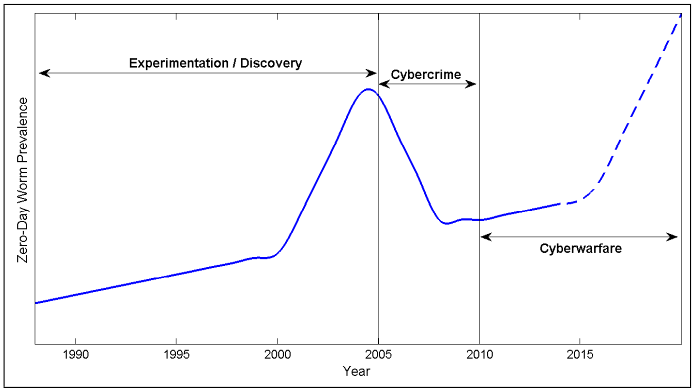


Fig. 1. Trend of Zero-Day Worm Prevalence

Although worms can be used in order to create botnets and carry out DDoS attacks, other methods have been chosen by attackers. Part of the reasoning for this, is that a large-scale, fast random-scanning worm outbreak is easily detectable, and it is often the intent of an attacker to avoid detection for as long as possible. Additionally, as has been shown by the Slammer outbreak [3], there is the possibility that a particularly fast worm can impede the network traffic, that in the case of a botnet, may disrupt the ability of an attacker to issue commands to or receive information from infected hosts.

As it has been shown by the worm-like Stuxnet outbreak [18], if it is the intention of an attacker to cause damage then the use of worms becomes a more attractive option. Although these have been isolated to targeted attacks to date, if it is the intention of an attack to disrupt communication or target the network infrastructure, such as in a cyberwarfare scenario, then the use of worms becomes a much more viable option. Additionally, if the motivation of an attack is to cause disruption of the Internet, then worms also present a viable option for attackers, even in the absence of a payload that causes damage.

Given the further shift of motivation from cybercrime to cyberwarfare, this also depends on the existence of wormable vulnerabilities, in order to exploit and carry worm attacks in the future.

V. Recent Wormable Vulnerabilities

Equation 3 presents a method of assessing whether or not a vulnerability is wormable. This Section presents five case studies of contemporary wormable vulnerabilities, along with their CVE code [10] for reference.

Microsoft Remote Desktop Protocol (RDP) - 13/03/2012 - CVE-2012-0002

The Microsoft RDP is a method for users to remotely access Windows-based hosts across a network. This vulnerability was present in a number of Windows versions, including XP, Vista, 7, Server 2003 and Server 2008. This allows an attacker to send a crafted packet on port 3389 to the host running RDP, and then potentially gain remote code execution. Having gained access to execute remote code, the attacker could then use this to send copies of the malicious packet

This wormable vulnerability is of particular note owing to the potentially large number of susceptible hosts to such an attack. W3Counter [21] reports that these recent editions of Windows constituted of approximately 3 billion Internet-connected hosts in 2012. As RDP is disabled by default, this requires being enabled manually. One estimate for the number of RDP enabled hosts is one in every 10,000 [22], or 300,000 hosts; resulting in a similar proportion of vulnerable hosts to the Code Red outbreak in 2001. As has been reported in two of the authors previous work [5], such a large proportion of susceptible hosts could result in a particularly virulent worm outbreak.

BigAnt Message Server - 09/01/2013 - CVE-2012-6275

The BigAnt instant messaging (IM) software is an instant messaging solution targeted towards business use. By using a buffer overflow present in the message server portion of the software, an attacker is able to send a crafted packet and execute remove code on the targeted machine. As the software links with Microsoft Active Directory, this can include ascertaining user account details, potentially having a wider impact than just the host running the message server. This can also lead to network access, allowing copies of the malicious packet to be sent to other hosts running the message server software.

Although lacking the install base of the Microsoft RDP vulnerability, this is of particular note owing to its use in a corporate setting, as well as potentially allowing access to further details that could lead to further issues. This vulnerability, as far as the authors are aware, also has yet to be patched and details of how to exploit this vulnerability are publicly available.

VMWare vCenter - 25/04/2013 - VMSA-2013-0006.1

VMWare vCenter is a management platform for virtualised hosts. A number of CVEs reported under the VMWare security advisory VMSA-2013-006.1 [23] detail how an attacker may leverage Microsoft Active Directory integration in order to gain authentication on Windows-based servers running vCenter (CVE-2013-3107), and then use this authentication in order to execute remote code using another vulnerability (CVE-2013-3079). This access provides administrative privileges to the host system, enabling the attacker to then send copies of the malicious packet/s used to other susceptible hosts.

As one of the largest vendors for virtualisation software, a vulnerability in VMWare software presents a scenario where a substantial number of hosts may be susceptible to an attack. Furthermore, access to the virtualisation environment may further allow access to the virtualised hosts that are currently running on it. This vulnerability has since been patched by VMware, however, it demonstrates that virtualisation can present a vulnerability for future worm outbreaks.

ASUS RT-AC66U Router - 26/07/2013 - CVE-2013-4659

The ASUS RT-AC66U router is a router produced for the consumer and small office market. Using a vulnerability in the Broadcom ACSD service allows an attacker to send a crafted packet on port 5916 causing a buffer overflow. This allows administrative access on the target device, providing remote code execution and the ability for the router to send copies of the malicious packet to other susceptible hosts. As far as the authors are aware, no known patch is available for this vulnerability and proof of concepts are currently available.

This vulnerability demonstrates that not only do server and desktop hosts require consideration when considering potential worm outbreaks, but also that of routing infrastructure. In addition to gaining access to further propagate itself, administrative access to the router may also allow for further attacks, including man-in-the-middle or denial of service attacks against hosts connecting to the Internet through this router.

systemd 208 and prior - 20/09/2013 - CVE-2013-4391

Designed specifically for Linux-based operating systems, systemd is a system management service, or daemon, that forms part of the Linux startup process. By using a crafted packet, a buffer overflow can be cause resulting in allowing remote code execution. In addition with another vulnerability, CVE-2013-4394 [10], administrative access can be gained, therefore allowing network access to send copies of the malicious packet/s to other susceptible hosts.

This vulnerability demonstrates that other operating systems, aside from Windows, can also be subject to a wormable vulnerability. It also demonstrates that software required by an operating system for basic functionality, as opposed to additional functionality in the case of the Microsoft RDP vulnerability, can also be vulnerable.

A. Host Discovery

As highlighted in the work reported by Shannon et al. [4] and Staniford et al. [12], the use of a hitlist is one key method of increasing the virulence of a worm outbreak. Given that a number of unpatched vulnerabilities have been highlighted, it is of note that there now exist a number of services that catalogue information provided through the use of meta-data. One such service, Shodan [24], is freely available and allows the collated download of search results at a small price. Such a service could be used in order to collate information prior to a worm outbreak, in order to create a hitlist.

B. Susceptible Population

A key factor in determining the virulence of a worm is the number of susceptible hosts that a worm can infect. As has been demonstrated in some of the authors previous work [5][13], and the measure of exploitability by Nazario et al. [7], the larger the proportion of susceptible hosts on a network both virulence and exploitability increase. In the case studies presented, those vulnerabilities that would provide the greatest number of susceptible hosts, are vulnerabilities in operating systems. Therefore, it is pertinent to further investigate operating system security.

VI. Operating System Security

A. Operating System Memory Security

The main method for exploited vulnerable hosts, allowing for remote code execution, has been the use of buffer overflow exploits (as demonstrated in table I). This has prompted the development of a number of techniques in order to prevent the writing of arbitrary data in the memory addresses that are being used by a program; and therefore providing remote code execution. The prevention techniques that are widely adopted in modern day operating systems are Address Space Layout Randomisation (ASLR), Data Execution Prevention (DEP), using No eXecution (NX) and canaries.

1) Address Space Layout Randomisation: ASLR is a countermeasure mechanism [25] adopted by operating systems to randomize the positions of executable code and data in memory at each run of a program. Randomising the base address of important memory structures, such as the stack and heap, makes the virtual address needed to perform a control-flow hijacking attack unknown. However, some techniques [26][27] have been reported that can bypass the randomness of ASLR mechanism.

a) Non-ASLR Memory: A non-ASLR module that runs on ASLR enabled operating system can be used to circumvent the ASLR protection mechanism. This can be a shared library in Microsoft Windows compiled without ASLR support for compatibility reasons. When an application that is non-ASLR is executed, the application tends to load its executables at runtime at a fixed memory address, thus allowing critical memory sections to be overwritten, or changing memory location. Additionally, using return-oriented programming techniques the contained data can be abused in order to leak additional memory addresses.

b) Information Disclosure: An information disclosure vulnerability can be used to leak memory locations of elements known to be at fixed addresses. For example, an out-of-bounds memory access vulnerability can be used to read a function pointer, and then send the value back to the remote server. Consequently, the server will control the size parameter of the function and accurately trigger an out-of-bounds read. As a result, the address of the public function is leaked. Based on this address, the memory layout of a corresponding executable file can be inferred.

c) Heap Spraying: Heap spraying is a technique used to allocate a substantially large amount of memory and fill it with a concatenation of multiple copies of a block of data. This is intended to create heap blocks using scripting languages so that a reliable location can be attained, then execute shellcode without looking for an offset in the memory address. This can greatly increase the probability that a chosen address will point to the beginning of the block even in the presence of randomisation.

2) Data Execution Prevention: Execution prevention [27][28] is another important countermeasure used to prevent arbitrary code execution even when an attacker has gained control over the processor's instruction pointer. This technique marks memory regions of executable application or service as writable or executable, but not both at a time. Popularly known as DEP on Microsoft Windows systems, it utilizes a hardware feature of the processor known as the NX bit. This marks writable memory regions, including the stack, as non-executable. Thus, when an address from this memory region is loaded as the instruction pointer, the processor will notice the non-executable flag and then raises a kernel level exception. The kernel will then send a segmentation fault signal to the program and thus terminate the program. Techniques used to circumvent DEP include return into libc, Return-Oriented Programming (ROP) and stack pivoting.

a) Return into libc: This technique [25] bypasses DEP by using the code of the running program or its shared libraries for malicious purposes instead of its intended use. This is achieved since the code is used by the running program itself, then the memory space utilised by the program is marked as executable. For example, in the Windows operating system an attack that uses WinExec and its functions (normally found in ntdll.dll) bypasses DEP as these are stored in an executable part of the memory. Thus malicious code can be copied to the executable memory space giving the attacker control of applications and services as described in [29].

b) Return-oriented Programming: This technique [25] allows an attacker to take control of the processor's instruction pointer and the stack area where return addresses are stored. Small pieces of code called gadgets are chained together to execute a chosen functionality instead of executing the intended functions. These gadgets are simple instructions followed by a return statement. For example, the statement in Figure 2 moves the content of the stack `esp` to `ecx` and then loads the next address from the top of the stack into the processor's instruction pointer through the return statement. This technique can successfully bypass DEP using WinExec as reported in [30].

c) Stack Pivoting: This technique [25] is an improvement of return-oriented programming by utilizing a special ROP gadget in order to make return-oriented programming possible through arbitrary overwrites. Having taken control of the processor's instruction pointer, an attacker will use the pointer to jump to a gadget that modifies the stack pointer to make it point to a controlled location. This can be accomplished directly through an arithmetic operation or by gadgets containing the `popq` instruction. It is intended that the controlled stack area will contain the ROP shellcode that will be executed subsequently.

3) Canaries: This is a compiler technique [31] that protects the stack by inserting a guard, a randomly chosen integer, at the start of the program between the protected region of the stack and the local buffers, i.e., a canary value is placed after the return address. Therefore, overwriting the return address will change the canary value, which is normally checked before a function uses the return address. The function will compare the value on the stack and the original value of the canary, if these values are different, then a message is generated in the system logs and the program will be terminated.

```
mov esp, ecx
ret
```

Fig. 2. Example Return-Oriented Programming Gadget

B. The Windows XP Opportunity

It has been estimated that Windows XP still constitutes 26% of all operating systems installed on desktop hosts [32]. As of the 8th April 2014, the extended support for Windows XP was discontinued. This meant that from this date there were no longer any security patches or support for this version of the operating system being made available for free. Although what is termed “critical patches” will be made available to paying customers. Additionally, after the 14th July 2015, the built-in anti-malware tools, Security Essentials and the Malicious Software Removal Toolkit, will no longer be supported.

Given this lack of support, if vulnerabilities are found in this version of the Windows operating system, it increases the likelihood that these systems will be susceptible to a future worm outbreak. This presents a particular issue, for instance, Slammer was able to cause disruption with less than 1% of the hosts at the time being susceptible to its infection vector [3], therefore it is reasonable that should a Windows XP vulnerability be exploited by a Slammer-like attack it could cause significant network disruption.

VII. CONCLUSION

Since the turn of the 21st century, zero-day worms have constituted a considerable threat to the Internet. Since 2005 there has been a reduction in the number of worm outbreaks, which can be attributed to a shift in the motivation of attackers from a period of experimentation and discovery to that of criminal activity. As such activity is better suited to the use of other types of malware, such as trojans, this reduction is reasonable. With the advent and increase in prevalence of cyberwarfare, worms once against become a weapon of choice for attackers, owing to their fast propagation and ability to cause considerable damage.

This paper explored the contemporary availability of wormable vulnerabilities and discusses the increased proportion of susceptible hosts made available by exploiting operating system vulnerabilities, highlighting the common techniques used in order bypass the most common techniques for preventing the exploitation methods used by zero-day worms. Furthermore, it highlights the opportunity that has arisen for attackers with the end of extended support, and future end of anti-malware support, for the Windows XP operating system.

REFERENCES

[1] B. Ediger, “Simulating Network Worms - NWS Network Worm Simulator,” http://www.stratigery.com/nws/, Sep. 2003, retrieved: 28th July 2014.

[2] C. C. Zou, W. Gong, and D. Towsley, “Code red worm propagation modeling and analysis,” in *Proceedings of the 9th ACM conference on Computer and communications security*. ACM, 2002, pp. 138–147.

[3] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, “The spread of the sapphire/slammer worm,” *IEEE Security and Privacy*, vol. 1, no. 4, pp. 33–39, 2003, retrieved: July, 2014.

[4] C. Shannon and D. Moore, “The spread of the witty worm,” *Security & Privacy, IEEE*, vol. 2, no. 4, pp. 46–50, 2004.

[5] L. Tidy, S. Woodhead, and J. Wetherall, “A large-scale zero-day worm simulator for cyber-epidemiological analysis,” vol. 3, no. 1. Universal Association of Computer and Electronics Engineers, 2013, pp. 69–73.

[6] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, “A taxonomy of computer worms,” in *Proceedings of the 2003 ACM workshop on Rapid malcode*. ACM, 2003, pp. 11–18.

[7] J. Nazario, T. Ptacek, and D. Song, “Wormability: A description for vulnerabilities,” *Arbor Networks (October 2004)*, 2004, retrieved: July, 2014.

[8] P. Li, M. Salour, and X. Su, “A survey of internet worm detection and containment,” *Communications Surveys & Tutorials, IEEE*, vol. 10, no. 1, pp. 20–35, 2008.

[9] C. Smith, A. Matrawy, S. Chow, and B. Abdelaziz, “Computer worms: Architectures, evasion strategies, and detection mechanisms,” *Journal of Information Assurance and Security*, vol. 4, pp. 69–83, 2008.

[10] M. Corporation. (2014, April) CVE - common vulnerabilities and exposures. Online. Retrieved: July, 2014. [Online]. Available: https://cve.mitre.org/

[11] N. Weaver, “Warhol Worms: The potential for very fast internet plagues,” http://www.iwar.org.uk/comsec/resources/worms/warhol-worm.htm, 15 Aug. 2001, retrieved: July, 2014.

[12] S. Staniford, D. Moore, V. Paxson, and N. Weaver, “The top speed of flash worms,” in *Proceedings of the 2004 ACM workshop on Rapid malcode*. ACM, 2004, pp. 33–42.

[13] L. Tidy, S. Woodhead, and J. Wetherall, “Simulation of zero-day worm epedimiology in the dynamic heterogeneous internet,” *Journal of Defense Modeling and Simulation*, 2013, in Press.

[14] E. H. Spafford, “The internet worm program: An analysis,” *ACM SIGCOMM Computer Communication Review*, vol. 19, no. 1, pp. 17–57, 1989.

[15] Panda Security. (2013, May) Pandalabs q1 report: Trojans account for 80malware infections, set new record. Online. Panda Security. Retrieved 28 July 2014. [Online]. Available: http://press.pandasecurity.com/news/pandalabs-q1-report-trojans-account-for-80-of-malware-infections-set-new-record/

[16] K. Stevens and D. Jackson, “Zeus banking trojan report,” *Atlanta, DELL Secureworks. http://www. secureworks. com/research/threats/zeus*, 2010, retrieved: July, 2014.

[17] C. Wilson, “Botnets, cybercrime, and cyberterrorism: Vulnerabilities and policy issues for congress.” DTIC Document, 2008.

[18] N. Falliere, L. O. Murchu, and E. Chien, “W32. stuxnet dossier,” Tech. Rep., 2011.

[19] S. Cherry, “How stuxnet is rewriting the cyberterrorism playbook,” *IEEE Spectrum. http://spectrum. ieee. org/podcast/telecom/security/how-stuxnet-is-rewriting-the-cyberterrorism-playbook*, 2012.

[20] J. Mirkovic and P. Reiher, “A taxonomy of ddos attack and ddos defense mechanisms,” *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 39–53, 2004.

[21] Awio Web Services LLC. (2012, November) W3counter - global web stats. Retrieved: July, 2014. [Online]. Available: http://www.w3counter.com/globalstats.php

[22] B. Krebs. (2012, October) Service sells access to fortune 500 firms. Online. Retrieved: July, 2014. [Online]. Available: https://krebsonsecurity.com/2012/10/service-sells-access-to-fortune-500-firms/

[23] VMWare Inc. (2013, October) Vmsa-2013-0006.1 vmware security updates for vcenter server. Online. VMWare Inc. Retrieved: July, 2014. [Online]. Available: https://www.vmware.com/security/advisories/VMSA-2013-0006

[24] D. Goldman, “Shodan: The scariest search engine on the internet,” *Webseite, Stand*, pp. 01–21, 2014.

[25] R. Hund, C. Willems, and T. Holz, “Practical timing side channel attacks against kernel space aslr,” in *Security and Privacy (SP), 2013 IEEE Symposium on*. IEEE, 2013, pp. 191–205.

[26] T. Wang, K. Lu, L. Lu, S. Chung, and W. Lee, “Jekyll on ios: when benign apps become evil,” in *Proceedings of the 22nd USENIX conference on Security*. USENIX Association, 2013, pp. 559–572.

[27] S. Röttger, “Malicious code execution prevention through function pointer protection,” 2013.

[28] A. Cugliari, L. Part, M. Graziano, and W. Part, “Smashing the stack in 2010,” *no. July*, pp. 1–73, 2010.

[29] N. Stojanovski, M. Gusev, D. Gligoroski, and S. Knapskog, “Bypassing data execution prevention on microsoftwindows xp sp2,” in *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*. IEEE, 2007, pp. 1222–1226.

[30] V. Katoch. Whitepaper on bypassing aslr/dep. Online. Secfence Technologies. Retrieved: July, 2014. [Online]. Available: http://www.exploit-db.com/wp-content/themes/exploit/docs/17914.pdf

[31] H. Marco-Gisbert and I. Ripoll, “Preventing brute force attacks against stack canary protection on networking servers,” in *Network Computing and Applications (NCA), 2013 12th IEEE International Symposium on*. IEEE, 2013, pp. 243–250.

[32] Net Applications. (2014, April) Desktop operating system market share. Online. Net Applications. Retrieved: July. 2014. [Online]. Available: http://www.netmarketshare.com/

Improving Resource Discovery and Query Routing in Peer-to-Peer Data Sharing Systems Using Gossip Style and ACO Algorithm

Hamdi Hanane, Benchikha Fouzia
LIRE laboratory, Department of Software Technology
and Information Systems, University of Constantine 2
Constantine, Algeria
emails: {Hamdihanane@hotmail.fr, f_benchikha@yahoo.fr}

***Abstract*—With the far-reaching significance of the Internet and the drastic advances in computer technology, more and more news and data are available on the Web. Peer-to-Peer (P2P) systems have become a popular way of sharing these data, and have drawn much attention of both academia and industry. The key and one of the most challenging design aspects in data sharing in P2P systems is how to find flexible, scalable and efficient mechanisms for searching and retrieving data. In the proposed approach, we use mobile agents to take advantage of a distributed system. To optimize the migration strategy of the mobile agents, we first resort to the biological behavior of ant colonies; the agents use trails leaved by other agents on peers they have visited. Then, to enhance the quality of the search results, reduce the randomness of peers and preserve their autonomy, we also introduce a social aspect, i.e., the friends list is used to gather peers having similar center of interest. To discover friends, the peers rely on the gossiping algorithm. We find that our contribution has three originalities distinguishing it from other approaches. The first one takes into account the two principal issues in data sharing in P2Pdatabase systems. The second one has the advantage to be totally independent of the centralized management. Finally, the third one is inherent to the resource discovery mechanism, which includes a social aspect using the friendship links.**

Keywords-P2P; Mobile agent; resource discovery; gossiping; ant colony optimization.

I. INTRODUCTION

With the far-reaching significance of the Internet and the drastic advances in computer technology, more and more news and data are available on the Web. Peer-to-peer systems have become a popular way of sharing these data, and have drawn much attention of both academia and industry. According to [1], P2P file transfer occupies 86.7% of the total file transfer traffic. The basic principle of P2P technology is that the peer acts as both a client and a server.

In the proposed research, we are interested in data sharing in P2P databases systems. Peer-to-Peer Data Management Systems (PDMS) have emerged recently; they combine P2P technology and distributed databases *Piazza* [2], *SomeWhere* [3] and *PeerDB* [4].

Although, coupling data integration techniques and P2P systems is efficient, it is essential to overcome some obstacles mainly due to the heterogeneity, decentralization and dynamic nature of P2P. In a P2P network, it is almost impossible to build or agree upon a mediation schema, with nodes joining and logging out continuously. Due to the absence of a centralized control, peers do not know the *a priori* location of the data they are looking for. The key and one of the most challenging design aspects here is to find flexible, scalable and efficient mechanisms for searching and retrieving data.

To take advantage of a distributed system, we must look for a distributed management solution. As the P2P networks, the paradigm of the mobile agent has been specifically developed for dynamic, distributed, open and heterogeneous environments. On behalf of network users, mobile agents execute software entities that are capable to migrate from one node to another in heterogeneous networks. Lange et al. [5] showed that mobile agents not only reduce the network load and overcome its latency, but also encapsulate protocols, execute asynchronously and autonomously, and dynamically adapt to changes.

Here, we rely on the mobile agents' intelligence and their capability of adapting their migration itinerary to the changing conditions. In that sense, in the proposed approach, we are inspired by the biological behavior of ant colonies. To optimize their migration strategy, mobile agents use trails leaved by other agents on peers they have visited. In addition, a mobile agent can make measurements anywhere on the network and take real-time decisions as well.

We also introduce a social aspect through the use of the friends list to gather peers having similar center of interest. This reduces the randomness of peers, which leads to better search results. It also enables robust self-monitoring and preserves the peer's autonomy. To discover their friends, the peers rely on the gossiping style; which proved to be very efficient for supporting dynamic and complex information exchange among distributed peers. They are useful for building and maintaining the network topology itself, as well as supporting a pervasive diffusion of the information injected into the network.

As the heterogeneity issue has been treated in a previous work [6]; in this paper, we mainly focus on the resource discovery and routing queries. The rest of paper is organized as follows. In Section 2, we first introduce a comparative study of the multiple research works then, to help understand the proposed approach, we present some of its relevant concepts. Next, in Sections 3, 4 and 5, we lay out the details inherent to the development of the approach. A summary of the results of the contribution along with the conclusions are given in Section 6.

II. Background And Related Works

Several authors focus on resolving the issue of data sharing in a peer-to-peer network. Research works [2][3][7] are interested in the resource discovery and the query routing issues. In a topology independent of central process, band failure and bottlenecks, ones' challenge is to offer a resource discovery method with a maximum accuracy. In this context, various techniques are proposed in the literature. They depend essentially on the network topology, such as distributed hash tables, centralized repository, semantic overlay networks (super peer) and flooding.

King [8] uses the Distributed Hash Tables (DHT) keep a tight control on the structured network and often find the desired information by means of the query for unique identifiers. However, this technique requires knowledge of the identifiers before any query is processed, and is too sensitive to failure. Moreover, each peer depends on other peers; which limits its autonomy. Indeed, the algorithm is based on the concept of successor, which may turn out to be faulty. Finally, this technique handles poorly keyword searches and complex queries.

The central repository, such as *BitTorrent* [9], relies on a centralized topology. A central repository is used to index all the peers and their respective data. This technique does not satisfy the distributed control criteria for P2P systems. It also creates a bottleneck, and limits scalability by concentrating all the resource information at a single point. In fact, if the server crashes, the whole network stops working.

Owing to its numerous advantages, the overlay semantic networks or the commonly known Super-Peer topology, used in *Piazza* [2] and *SenPeer* [7], has become very popular. This topology has though a major drawback, namely, its mishandling of client/server at a group level. For instance, if for any raison, the Super-Peer fails, all peers turn out to be unavailable to the assignment of a new Super-Peer. Furthermore, the queries are sent only to related semantic groups. In this case, groups that do not have any semantic relationship with the query are simply ignored.

The flooding technique *Gnutella* [10] uses a recursive process to send a or query to all nodes in the network. Thus, it does an exhaustive search. To avoid messages travelling indefinitely in the network, the number of nodes a message can visit is set to a limit called Time To Live (TTL) [10]. PeerDB [4] proposes a different approach for the flooding technique. In this case, mobile agents are used for query routing and processing as well.

Other works in the literature use gossip style for information dissemination in the network. In *TRibler* [11], a P2P television recommender system is proposed. The authors suggest a social networking system, built on a P2P network which is *Bittorrent-based* file-sharing client [9]. To establish friendship links between the source and the target peers, this system relies on a gossip style. More P2P Recommendation Systems *are found in P2Prec* [12] and *P2PREcommender* [13].

Further research works, such as [3][9], are interested in schema mediation; omitting the resource discovery and the query routing aspects. Most of the peer-to-peer systems only deal with non structured or semi-structured data *Gnutella* [10]. However, some search approaches allow sharing structured data as well as doing searches based on its content. Most of these works are based on ontology to address the heterogeneity issue, such as global ontology [8][6] and local ontology approaches [3].

Before introducing the proposed approach in Section 3, let us first get familiar with some of its relevant concepts, i.e., P2P-database, Ant Colony Optimization algorithm (ACO) and gossiping protocol (also known as the epidemic protocol).

A. Peer-to-Peer databases systems

A P2P database system (PDBS) is perceived as a collection of autonomous local repositories which interact in a peer-to-peer style [2]. On the other hand, a Peer Data Management System (PDMS) is a triplet-set, i.e., S = <P; S; Mi> where, P, S and M are sets, respectively of autonomous peers, heterogeneous schemes and schema mappings; each of which enables the reformulation of queries between a given pairs of schemas [15]. A PDMS is a distributed data integration system providing transparent access to heterogeneous databases without resorting to a centralized logical schema. Instead of imposing a uniform query interface over a mediated schema, PDMSs let peers define their own mappings, supplied locally with regards to different schemas pairs or groups of pairs, to be used to reformulate queries.

B. Ant Colony Optimization (ACO)

The Ant Colony Optimization (ACO) [14] algorithm is inspired by the behavior of ants colonies. That is, the ants mark their trails through different intensities of pheromones, i.e., secreted chemical substances to communicate between them. In doing this, they create a system in which the best route, chosen by the others ants, would correspond to the highest level of pheromones. An important characteristic of this algorithm is to dynamically absorb changes in the graph. This makes this characteristic practical in dynamic network routing systems and also suited to P2P networks [14].

C. Gossiping

Gossiping also known as *epidemic* protocols were first introduced in 1987 by Demers et al. [16], who employed them in propagating updates in loosely replicated databases. These protocols have the effect of maintaining mutual consistency among the replicas. Gossiping protocols have mostly been associated with dissemination of information [17]

By omitting specific details, gossiping protocols work through a simple model. Each node has a complete view of the network, and periodically picks a random node from the whole network to exchange data with. This allows information, known by any node, to spread to the whole network with very high probability [17].

III. PROPOSED APPROACH

As a solution to the two principal issues for data sharing in P2P network, we introduce, in this section, the proposed approach.

To address the heterogeneity issue, the proposed approach relies on domain ontology. This solution is easy to implement. It takes into consideration the P2P characteristics, namely, autonomy, scalability and decentralization. In addition, to the advantages of semantics in data integration, the ontology provides users a common vocabulary to ensure unambiguous communication between heterogeneous sources. Each peer has a copy of the domain ontology. When joining the network, the peer must provide mapping between its local schema and the domain ontology. These mappings are stored locally and used for subsequent queries reformulation.

As stipulated in Section 1, here we are not concerned with providing solutions to the heterogeneity issue. We rather focus on the resource discovery and query routing issue. In doing this, we propose an agent-based architecture which obeys the P2P criteria. We, then, propose a mechanism of resource discovery and query routing, which takes advantage of using mobile agents and optimizes their migration strategy by avoiding an excess of message traffic through the use of the ACO. Finally, introduce a social aspect using the friends list to group together peers having similar center of interest. This reduces the randomness of peers, which leads to better search results. It also enables robust self-monitoring and preserves the peer's autonomy. To build and manage this friends list, we use the gossip style.

IV. AGENT-BASED ARCHITECTURE

Our solution is a distributed system evolved in a dynamic environment with peers joining and leaving the system.

When it comes to designing this type of system, agent technology is suitable, because multi agent systems not only allow the sharing or distribution of knowledge, but also the achievement of a common goal.

A. Mobile agent technology

In a broad sense, an *agent* is any program that acts on behalf of a (human) user. Then, a *mobile agent* is a program which represents a user in a computer network. To perform some computation on user's behalf, this program is capable of migrating *autonomously* from one node to another. In order to make possible previous difficult fault tolerance and distribution heuristics, agents are also able to perceive their environment and communicate with other agents [18].

Using agents in P2P networks offers many advantages.

- Reduction of the use of bandwidth and communication costs.
- Ability of adapting to a dynamical P2P environment.
- Asynchronous execution and fault tolerance.
- Autonomy.
- Cloning and dispatching of a mobile agent in different directions.

B. Overview of the architecture of a peer

As shown in Figure 1, a PDMS, as defined in Section 2.A, consists of a network of nodes referred to as peers.

Since PDMSs are a mechanism of a decentralized sharing of data, mappings are not controlled in any central manner. The only assumption we can make is that any peer that joins the system provides some mappings between its local schema and the domain ontology. Figure 1 shows an overview of the architecture of a peer. It is formed by different parts, which are succinctly described as follows.

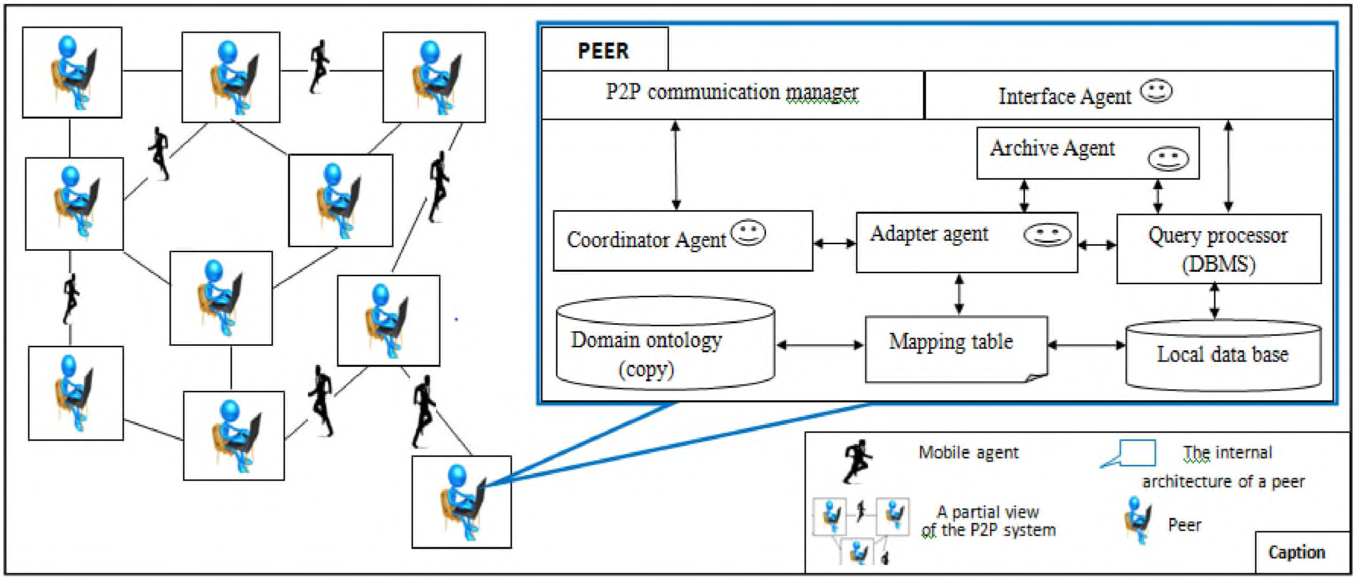


Figure 1. Overview of the proposed architecture.

- P2P network: Our approach relies on a pure unstructured P2P topology. The network is composed of a set of peers interconnected in an unstructured manner.
- Peers or users of the network: A set of peers participating in data sharing. Each peer has a partial view of the network, and plays the role of client, server and mediator.
- Mobile agents: The mobile agent travels over the network to meet the requirements of its initiator peer.

Here, the general architecture of a peer, shown in Figure 1, satisfies not only the P2P characteristic, but also the requirements of the proposed approach.

A. *The internal structure of the Peer*

In a P2P network, each peer must be able to play the role of-data provider, data requester and mediator.

To fulfill these roles, each peer must contain at least a:

- Data source,
- Copy of the domain,
- Set of mappings between the elements of the domain ontology and the local schema.
- Set of agents responsible of information retrieval and data mediation.

According to the proposed internal architecture of a peer, Figure 1, there are five types of agents.

1) Interface Agent (IA): It is a stationary agent that receives all user queries and process them locally. If necessary, the *IA* sends the processed queries to the coordinator agent. When the *IA* receives results, it presents them to users.

2) Coordinator Agent (CA): This agent coordinates and manages the other agent's work. It also creates mobile agents and dispatches them over the network. The social behavior of the peer is also managed by this agent; it is responsible of the creation and the updating of the friends list. Figure 2 shows the internal structure of the *CA*.

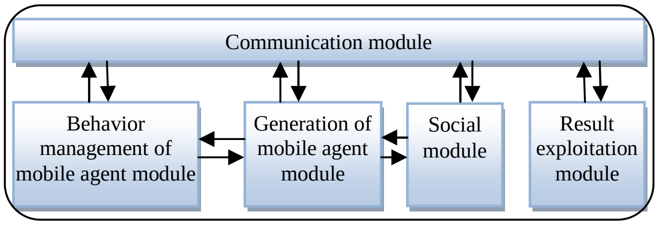

Communication module
Behavior management of mobile agent module
Generation of mobile agent module
Social module
Result exploitation module

Figure 2. Coordinator agent.

3) Social Mobile Agent (*SMA*): This agent is created by the *CA*. To collect data, the *SMA* travels over the network. Figure 3 shows the proposed *SMA* architecture. It contains

- A communication module to communicate with its *CA*, any peer it visits, and any other agents it may meet.
- A mobility manager and a cloning manager module. The social module is responsible of the social behavior of the mobile agent; it holds the interest list of its Creator Peer (CP) and measures its matching inherent to the interest list of other agents it may meet.
- A result exploitation module which is responsible of the query execution results.

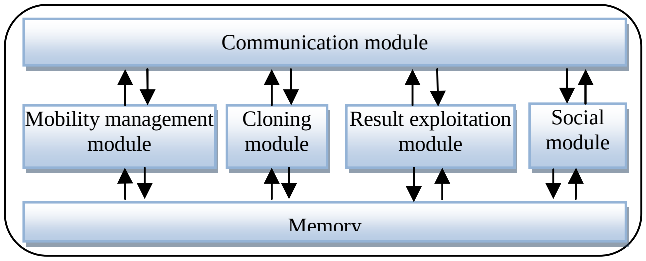


Figure 3. Social mobile agent.

4) Adapter Agent (AdA): This agent translates the user's request from the local language (local schema) to the global language (domain ontology), and vice versa. To do so, it uses the mapping tables created by the peer when it joins the network for the first time.

5) Archive Agent (ArA): This agent stores the resolved queries at time T, and checks its memory for each new peers request, to know whether or not this request has been already resolved earlier. It also memorizes the trails leaved by the mobile agents which have visited this peer.

B. *Functioning Principle*

1) Basic steps

To better understand the proposed approach, we consider a motivating example of an application for data sharing in a P2P network. We are interested in data sharing in the field of medical research; therein, it is mandatory to share data between different research institutes, doctors and scientists. To this end, let us consider a network of scientists interested in sharing the locally stored data. The scientists typically store different kinds of data, including papers and reports they have written, articles and theses they have downloaded, information about conferences, seminars and other scientific events, reports about patients they have treated, etc. The goal of such a network is that each user (peer) can discover data, provided by other peers (users) and sends queries to them, accordingly. For instance, to exchange reliable information with one another, one may want to know which peer has already treated a patient who has lived with Alzheimer's disease for 10 years. Likewise, another peer may also want to look for datasets used by others for some kinds of experiments. In this last scenario, each scientist should have its own database. In order to overcome heterogeneous schemas, databases must undergo an integration to allow scientists access data from other researchers in a transparent manner, irrespective of problems of heterogeneity or distribution.

To this effect, let us assume that a new researcher noted Pi joins the network, and would like to make a research. Here are the basic global steps of the proposed mechanism; from Pi joining the network until the acknowledgement of a query. The details of these steps are left for Section 5.

- Pi, first builds the mappings table between its local schema and the domain ontology. These mappings are stored locally to be used later for query reformulation.
- Next, he builds its friends list from its neighbor list which will be detailed in Section 5.1
- Then, he makes, on his local schema, a query about the search he would like to do. Only then, the query is translated into the vocabulary of the domain ontology.
- Pi creates mobile agents and sends them along with the query over the network.
- When a mobile agent arrives at a peer Pj, it transmits it the query, which prior to running it, translates it into the vocabulary of its local schema.
- The mobile agents come back to their initiator peer, Pi along with the data and the information they have collected.

2) Mobile agent life cycle

The life cycle of a mobile agent, Figure 4, designates all progression steps of the agent; from its creation until its death. The mobile agent creation can be initiated either by a user agent or by another mobile agent. Figure 4 illustrates the life cycle of our mobile agent.

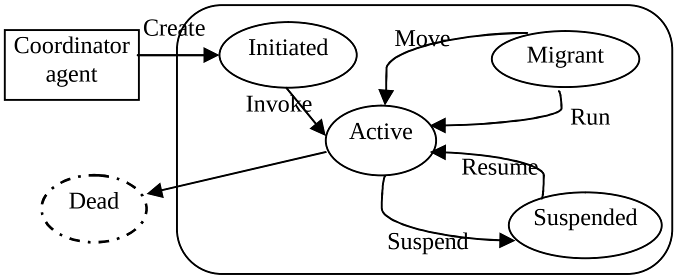


Figure 4. Mobile agent life cycle.

Initiated: The agent has been created but has not registered yet. As it has neither a name nor an address, it cannot communicate with other agents.
Active: The agent has been registered and has a name. In this state, it can communicate with other agents.
Suspended: The agent is stopped because its thread is suspended.
Migrant: The agent is moving to a new location.

Dead: The agent has finished and his thread has ended his execution and it is not any more in the AMS.

V. Gossiping And Aco Based Resource Discovery

As explained earlier, resource discovery and query routing are crucial issues for data sharing inP2P networks. Using mobile agents, in this section, we propose a mechanism of resource discovery and routing queries that couples the gossiping with the ant colony optimization.

Our mechanism is divided into two parts.

i. The first one deals with building and managing the friends list. That is, we exploit the gossiping to discover and group together users having similar interests "Friends".
ii. The second one deals with guiding the agents in their migration phase. Using the links established so far with the ACO, we optimize the routing strategy of the mobile agents.

This mechanism reduces the randomness of peers, which leads to better search results. It also enables robust self-monitoring and preserves the peer's autonomy. In such a way, it eases information dissemination, as peers discover new content and new peers. Finally, it is robust, resilient to failure and easy to implement.

a) Building and managing the friend list

Our work differs from the conventional gossiping model through the following points.

First, we leave out the assumption of complete knowledge of the network. Each node has only a partial view of the network.

Then, the type of data is such that nodes exchange their list of interest.

After a gossip exchange, nodes update their friends list to incorporate a part of the received links. In doing this and continuously refreshing their friends list, nodes can self-organize into better and suitable topologies.

Basically, our gossip protocol proceeds as follows. A peer Pi knows a group of other peers or contacts, which are maintained in a list called Pi's view. Periodically, with a gossip period noted Tgossip, Pi picks a random neighbor Pj from its view to gossip. Then, peers Pi and Pj exchange information. The gossip algorithm we propose here is inspired by gossip-based approaches for P2P membership management [17].

Each peer has a partial view of the network. It knows a small but continuously changing set of other peers. This view is divided into two lists. While the first one is the list of neighbors assigned by the system using bootstrapping, i.e., neighbors list; the second is the list of peers having in common the same centers of interest, i.e., friends list.

Note that each peer also maintains a list of interest. This list contains keywords that represent the issues and research topics for which the pair is most interested in. These keywords are annotated to the domain ontology.

Our friend discovery protocol is guided by the Buddycast algorithm described in *Tribler* [11]. To discover a friendship link, each peer Pi, in our algorithm, periodically sends a Gossip message to its neighbors, and updates its current friends list, accordingly. When a peer Pj receives a Gossip message, i.e., a message containing the IP (Internet Protocol) address, port of the Pi and its list of interest, it retrieves the list of interest and compares it to its interest list. If the lists are similar, it adds it to its friends list, and replies with a friendship invitation. Otherwise, it replies with a message saying, 'we are not interested in the same topics'.

Peers are conceived such that no other contact with the same peer can be made for the next three months. That is, we suppose that a researcher does not frequently change its research topics.

Two threads are defined in Figure 5 and Figure 6. They describe the gossip behavior of each peer Pi. The first, called 'active behavior', is executed every Tgossip time unit. It describes how Pi initiates a periodic gossip exchange. The second, called ‘passive behavior’, shows how it reacts to a gossip exchange initiated by some other peer.

```
Algorithm 1 Gossip behavior of user u
// active behavior, runs periodically every T time units
Let N(P) be the set of actual neighbors
Let F(P) be the set of actual friends.
Loop
(wait Tgossip)
  For all Pi □ N(P) do
   NewPeers← neighbors.SelectRandom ()
      For all P□ NewPeers do
 if P □/ F(P) then
          Connect with P
          GossipMsg = (myAddress, myInterestList)
          Send GossipMsg to P
           Receive GossipMsg from p
         endif
      end for
  endfor
end loop
```

Figure 5. Active behavior of a peer.

```
Algorithm 2 Gossip behavior of pr
//passive behavior Runs when contacted by some peer
Forever do
waitGossipMessage()
Receive GossipMsg from U
GossipResp = (myAdress, reponse)
Send GossipResp to U
```

Figure 6. Passive behavior of a peer.

b) Adaptive migration strategy

Adaptive migration strategy is a way of guiding an agent during its migration step. Our migration model is based on direct and indirect cooperation between agents. Direct cooperation takes place when the agents communicate with each other~~,~~ whereas, indirect cooperation exploits the notion of mobile agent’s trails. The idea of the indirect cooperation mechanism is based on the stigmergy theory. This ant colony-based theory was first developed by Grassé [19]. Explicitly, it uses the fact that ants do not communicate directly with each other.

Once a query is formulated in the proposed model, the created *SMA* dos not receive any predefined path; it builds its itinerary by itself as and when it travels over the network. To achieve this, we resort to the Ant Colony Optimization algorithm, in which an agent uses trails left by other agents on peers they have visited. Using this method, our approach enhanced the agent’s autonomy.

- ***Migration algorithm***

The migration algorithm is used to show how the mobile agents would behave while searching for information in the network. It is described as follows.

- If Pi desires to make a search in the network, it first formulates a query, and then runs it locally to find out whether or not the sought-after data exist in the local database. In the affirmative, the search stops. Otherwise, the query is reformulated by the *AdA* in the domain ontology vocabulary, and transmitted to the *ArA* to check if the query has already been resolved. In the affirmative, the query is transmitted to the *CA* along with the address of peers that have answered. Only then, the *CA* creates and sends mobile agents to the peers concerned.
- Otherwise, a social mobile agent is created by the Pi. This peer is known as the *CP* of this particular mobile agent as well as all of its clones. The mobile agent is created by the *CA* of the *CP*. The *SMA* is given in the form of parameters; i.e., information about the *CP* (name, address and port), energy, which corresponds to the maximum number of peers that may be visited before it must return to its CP, and a cloning factor, used to determine how many times the mobile agent may be cloned at any peer. Note that the energy and the cloning factor parameters control the depth and breadth of the search. Therefore, they handle the maximum number of peers that may be visited and the friends list as well.
- The SMA clones itself as many times as the number of friends present in the list it received. This allows a mobile agent to be sent to each of these friends. Note that if a destination cannot be located, the SMA tries the next one in the list.
- Upon arrival at each peer, the *SMA* checks whether or not it or its replica has already visited this peer. The details of how agents exploit the trails are given in Section 5.2.c In the affirmative, the *SMA* stops its migration and comes back to the *CP*. Otherwise, it decrements its energy; i.e., E←E-1, updates its migration history and that of the Host Peer (*HP*), with information about the *CP*, and leaves a trail of its visit. It also updates itself with information about the current peer.
- Then, the *SMA* gives the query to the *CA* of the *HP*, which in turn transmits it to the *AdA* of the *HP*, to be reformulated in the local schema vocabulary. Only then, the *CA* runs it and search for the result in the *HP* database, to be transmitted to the *SMA*, which stores it and then moves to the next destination.
- When the mobile agents’ energy has dissipated, and there are no more destination to visit, they then go back home.
- Once at home, the agent and its clones update the *CP* with the information they have so far collected throughout their journey and then dispose of themselves.
- During the course of its journey, the peer’s machine may shut down while the *SMA* is still collecting

information. The agent has then the capacity to wait for the peer's return. In such a case, it goes to sleep and does not wake up until the machine is powered on again.

- ***Ant Colony Optimization***

Without trails, the agent chooses a random schema migration. In the proposed approach, a trail designates the on sight visit tracking of an agent, or its next destination. This may lead to less information being acquired about the network. However, it keeps the information up to date, and yet prevents peers that form cycles from having to deal repeatedly with mobile agents from the same *CP*.
Social Mobile Agents use trails as follows.

- If two SMA, from the same *CP*, arrive at the *HP* within a preset time period, the last agent arriving verifies its migration history.
  - If the latter is empty, therefore E=E-1. It means that the current *HP* is the agent's first destination. Then, it sends a message to its CP, and suicides.
  - Otherwise, it stops its migration and returns to the *CP* along with the information it has already collected.
- Upon arrival at a node, the *SMA* asks the archive agent of the actual *HP* whether or not a replica of itself has visited this place.
- It also asks the archive agent about the trails of agents that have visited that peer in the last T time, and verifies if there are trails of an agent that have been created by a friend peer.

If there is one then, the *SMA* clones itself and sends the clone to the destination of that agent (we suppose that a friend of friend is potentially a friend). Upon arrival, the clone compares its interest list to the host peer's. If there is a correspondence, it runs, and comes back to its *CP* with the results of the query and the information about a new friend it has made. Otherwise, it destroys itself.

- If an agent meets another agent on sight, they exchange their interest lists. If there is a correspondence, they exchange their *CP* addresses and their future destinations.

VI. Conclusion And Future Work

In this paper, we have tackled the problems of resource discovery and query routing in P2P data sharing systems. In doing this, we have first studied and analyzed the existing approaches and mechanisms found in the literature. Then, we have proposed a mechanism for resource discovery and query routing. For this purpose, we have used an agent-based architecture, a resource discovery and a query routing mechanism. This latter is two-fold; one relies upon a gossip protocol to build and manage the friends list, and the other on the ACO to optimize the mobile agent migration.

As we have shown throughout this paper, our contribution exhibits three originalities. The first one takes into account the two principal issues in data sharing in P2Pdatabase systems. That is, while other PDBMS treats just one issue, the proposed one handles both hiding the heterogeneity of data from different sources through domain ontology, and developing an efficient resource discovery mechanism. The second one has the advantage to be totally independent of the centralized management; hence, the peers are completely autonomous for both schema mediation and resource discovery. Finally, the third one is inherent to the resource discovery mechanism, which includes a social aspect using the friendship links. This reduces the randomness of the peers, leading to better search results. It also takes advantage of making use of mobile agents in P2P networks, as seen in Section 4.1, and optimizes their migration through ACO features. In a future work, we will focus on conducting more experiments on the proposed mechanism by means of the PeerSim simulator [20].

References

[1] "Cisco Systems, Inc, Cisco Visual Networking Index: Forecast and Methodology, 2012-2016,"http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ ns537/ns705/ns827/white paper c11-481360.pdf.

[2] A. Halevy, Z. Ives, J. Madhavan, P. Mork, D. Suciu, and I. Tatarinov. "The piazza peer data management system". IEEE Transactions on Knowledge and Data Engineering, 16(7):787–798, 2004.

[3] P. Adjiman, P. Chatalic, F. Goasdoué, M.C. Rousset, and L. Simon."Somewhere in the semantic web". In PPSWR 2005: International Workshop on Principles and Practice of Semantic Web Reasoning, 2005, pp. 1–16.

[4] W. S. Ng, "Adaptive p2p platform for data sharing", PhD Dissertation, National University of Singapore republic of Singapore, March, 2004.

[5] G. Karjoth, D. Lange, and M. Oshima, "a security model for aglets," IEEE Internet Computing, 1997, vol. 1, no. 4, pp. 68-77.

[6] H. Hamdi, F. Benchikha "An agent and ontology based approach for data mediation in P2P ", International Conference on Artificial Intelligence and Information Technology, Ouargla, Alegria, 2014.

[7] D. C. Faye "Semantic data mediation in SenPeer, A peer data management system", PhD Dissertation, Nantes University, 2007.

[8] M. R. A. King, " Locating data sources and query optimization in distributed peer-to-peer environment", PhD Dissertation, Toulouse III – Paul Sabatier University 2010.

[9] J. A. Pouwelse, P. Garbacki, D. H. J. Epema, and H. J. Sips. "The Bittorrent P2P_le-sharing system : Measurements and analysis". In 4th Int'l Workshop on Peer- to-Peer Systems IPTPS, vol. 3640, 2005, pp. 205-216.

[10] Gnutella http://rfc-gnutella.sourceforge.net/index.html, retrieved : June, 2014.

[11] J. A. Pouwelse, J. Yang, M. Meulpolder, D. H. J. Epema, and H. J. Sips "BuddyCast: an operational peer-to-peer epidemic protocol stack" in Delft University of Technology Parallel and Distributed Systems Report Series 2008.

[12] R. Akbarinia, E. Castanier, R. Coletta, F. Draidi, E. Pacitti,and D. Parigot "P2Prec: a social-based p2p recommendation system" in Proc: the 20th ACM Conference on Information and Knowledge Management, (CIKM2011), United Kingdom, 2011, pp. 2593-2596.

[13] R. Baraglia, M. Mordacchini, P. Dazzi, and L. Ricci "A P2P REcommender system based on Gossip Overlays (PREGO)" 2010 10th IEEE International Conference on Computer and

Information Technology (CIT 2010), West Yorkshire, UK, 2010, pp. 83 - 90.

[14] M. Dorigo and C. Blum. “Ant colony optimization theory: a survey”. Theoretical Computer Science, Elsevier Science Publishers Ltd., Essex, UK, 2005, vol. 344, n. 2-3, pp. 243-278.

[15] P. Cudre-Mauroux “Peer data management system” MIT Computer Science and Artificial Intelligence Laboratory Cambridge, MA, USA, 2007.

[16] A. Demers, D.Greene, C.Hauser, W.Irish, J.Larson, S.Shenker, and al. “Epidemic algorithms for replicated database maintenance”.In Sixth Symp. on Principles of Distributed Computing, 1987, pp. 1–12, New York, NY, USA. ACM Press.

[17] S. Voulgaris “Epidemic-based self-organization in peer-to-peer systems” PHD Dissertation, University of Amsterdam, 2006.

[18] J. Nair “An application for resource discovery in a peer to peer network using mobile agents” in International Journal of Emerging Technology and Advanced Engineering, March, 2012, vol. 2, Issue 3.

[19] P. P. Grassé. “The reconstruction of the nest and inter-individual coordination among Belicositermes natalensis and Cubitermes sp. the theory of stigmergy: Test interpretation of the behavior of termites manufacturers. Social insects”, 6 :41–80, 1959. 106.

[20] Official website of the PeerSim simulator, http://peersim.sourceforge.net/, retrieved: June, 2014.

Improving Network Traffic Anomaly Detection for Cloud Computing Services

Ana Cristina Oliveira*†, Marco Spohn†‡, Reinaldo Gomes†, Do Le Quoc§ and Breno Jacinto Duarte¶
*Research Group on Convergent Networks (GPRC) - Federal Institute of Paraíba (IFPB)
Campina Grande, Paraíba, Brazil
†Systems and Computing Department (DSC) - Federal University of Campina Grande (UFCG)
Campina Grande, Paraíba, Brazil
‡Federal University of Fronteira Sul (UFFS)
Chapecó, Santa Catarina, Brazil
§Systems Engineering Group (SE Group) - Technical University of Dresden (TU Dresden)
Dresden, Germany
¶Research Group on Usable Security and Ubiquitous Communications - Federal Institute of Alagoas (IFAL)
Maceió, Alagoas, Brazil
Emails: ana.oliveira@ifpb.edu.br,{maspohn,reinaldo}@dsc.ufcg.edu.br,
do@se.inf.tu-dresden.de,brenojac@ifal.edu.br

Abstract—**Efficient network traffic anomaly detection is a widely studied problem on avoiding attacks and unwanted use of communication infrastructures. Existing techniques to detect, prevent or monitor these attacks are usually based on known thresholds, on the construction of profiles of normal traffic patterns, or on signature pattern matching of anomalous behavior (i.e., viruses and attacks). On the other hand, there are dynamic techniques that strive to predict the system's clutter degree; i.e., the system entropy, supposing that outliers translate to anomalies. We have developed and analyzed the accuracy of a network anomaly detector for Cloud Computing Systems based on the entropy of network traffic metrics. Although entropy-based solutions do not suppose hard knowledge of the system, the results point out to the need for more accurate adjustment of system parameters, taking into consideration the nature of the data, frequency of events, and the variation of metric values. To improve the results, unsupervised machine learning algorithms were added to the anomaly detection process.**

Keywords–Network traffic anomaly detection; Cloud Computing; Entropy; Machine learning.

I. Introduction

Network traffic analysis in cloud environments is one of the most important tasks in cloud management to guarantee the quality of services, validate performance of new applications and services, build accurate network models and detect anomalies in the cloud. The network traffic produced by cloud computing systems reveals users' behavior regarding service utilization, once all services are accessed via the network. Traffic analysis and the recognition of all significant application flows are important tools for modeling service usage, building up patterns for identifying normal system operations [1].

Additionally, network communication between cloud provider and its customers affects significantly the performance of most cloud-based applications [2]. Analyzing network traffic will provide insights on the performance and behavior of application and services deployed in clouds. Therefore, it is necessary to develop network traffic measurement and analysis techniques to improve availability, performance and security in cloud computing environments.

On the other hand, managing and analyzing network traffic of large scale cloud systems is a challenging task. The techniques used to monitor and analyze traffic in conventional distributed systems differ from cloud computing systems. In conventional approaches, assumptions are made that network flows follow some patterns, which is acceptable for corporate applications, but cloud applications may have significant changes in traffic patterns [3].

The term **anomaly** is fairly generic and it covers attacks, unwanted traffic on the network due to misbehaving applications, packet loss, and undesired traffic injection from not allowed applications. In this context, one question is raised: can we actually have a monitoring system able to detect any sort of network anomaly without looking specifically for it? There are works that proposed the idea of identifying any type of anomaly by monitoring metrics' entropy [4][5][6]; i.e., analyzing the behavior of an application by monitoring the degree of concentration or dispersion of the target metrics' distribution.

In this paper, we focus on investigating techniques to detect anomalies within cloud computing network traffic. We adapted and implemented the Entropy-based Anomaly Testing (EbAT) methodology into the context of cloud computing network traffic monitoring [7]. We strived to identify if EbAT is suitable for monitoring cloud computing systems. Considering it is a scalable and lightweight technique, which is a non functional requirement for cloud computing systems, since they strongly depend on the network support in large scale.

We concluded that the EbAT technique by itself can be improved to monitor network traffic, especially for dynamic systems that may change the network load quickly. To improve the anomaly detection accuracy for cloud computing network traffic, we developed a new lightweight approach based on the EbAT method and unsupervised machine learning anomaly detection.

The contributions of this work are fourfold: (i) implementation and analysis of the EbAT technique applied to cloud computing network traffic; (ii) feasibility analysis of the entropy-based method to anomaly detection of cloud computing network traffic; (iii) implementation and analysis of unsupervised machine learning technique for anomaly detection; (iv) proposal and implementation of a novel lightweight method to improve network traffic anomaly detection for cloud computing systems.

The remainder of this paper is organized as follows. Related work is presented in Section II. The entropy-based and the machine learning anomaly detection methods are described in Section III. The novel approach is proposed in Section IV. The validity of the proposed technique is addressed in Section V. To conclude, Section VI contemplates final remarks.

II. Related Work

Wang [7] proposed a method for online generic anomaly detection based on the entropy of any sort of metric distribution or composition of metrics. Such technique is called EbAT. The results are achieved by establishing entropy time series resulting from visual spike detection, or wavelet analysis, instead of the observation of individual thresholds for the metrics. However, it assumes that some parameters are statistically estimated.

Wang *et al.* [5] conducted an experiment to demonstrate the feasibility and accuracy of the EbAT method in comparison with threshold-based anomaly detection procedures. They injected faulty operations at the application level, which were analyzed using CPU and memory metrics, and correlation between read and write operations at virtual disks.

Wang *et al.* [6] compared the EbAT technique and Gaussian model for anomaly detection of system metrics (e.g., CPU and memory). According to their study, the Gaussian model presented lower values for the *recall* metric. Notwithstanding, we believe those techniques may be applied together to support the anomaly detection decisions.

Benetazzo *et al.* [8] proposed the analysis of aggregate traffic by determining empirical rate-interval curves (RICs), which consist of dividing the flow measurements in quantiles, striving to delineate scaling properties and other metrological diagnostics. The RIC-based method characterizes network traffic without requiring a priori knowledge of the underlying flow model; however, the proposed method is quite costly.

Nychis *et al.* [9] analyzed the anomaly detection ability of different entropy-based metrics. They pointed out that the port and address distributions are strongly correlated to the detection capacity. In other words, both metrics provide similar results when detecting network traffic anomalies. In addition, the metrics have limited utility in detecting port scan attacks and flood attacks. The authors also found that behavioral metrics are less correlated with other metrics. However, their work was applied in a university network backbone; thus, the characteristics of the metrics would be different from a cloud computing environment.

Quan *et al.* [10] compared two entropy methods, network entropy and normalized relative network entropy (NRNE) to classify network behaviors. Two different probability distributions could share the same entropy value, even having discrepant probability vectors, and it is a problem regarding the technique. To avoid this problem, the authors employed the concept of relative entropy, or Kullback-Leibler (KL) deviation, which represents the difference between two probability distributions. The NRNE performed better; on the counterpart, it demands more input attributes to detect abnormal network behaviour.

Smith *et al.* [11] proposed an autonomic mechanism for detecting anomalies in cloud computing systems similar to EbAT. The authors defined a set of techniques involving data transformation to standardize the data format for analysis, an extraction phase to reduce data size, and unsupervised learning using clustering algorithms to detect which nodes are behaving in a different manner from the others (outliers). The anomalies are computed based on the system behavior.

III. Fundamentals

A. Entropy-Based Anomaly Detection

We may divide the EbAT [5][7] technique in three steps: (a) metric collection; (b) construction of entropy time series; and (c) processing of entropy time series. This technique is metric-independent, i.e., we may collect and analyze the most important metrics related to network traffic, or to application-level performance, for instance. Those steps are described in the following sections.

1) Construction of Entropy Time Series: The metrics being analyzed at a moment are placed into a look-back window of size n, where n means the number of samples for the metric (or metrics) that will take part of the analysis. That window slides as the new metrics are being produced.

Note that multiple types of metrics may be monitored and analysed altogether. Before constructing the entropy time series, the data is pre-processed. This pre-processing consists of normalizing and binning the data. Those phases are characterized as follows:

Data Normalization Phase: it consists of dividing all sample values in the current look-back window by the mean of all values of the same type that belong to the window in question.

$$s_{i,j}^{'} = \frac{s_{i,j}}{\frac{1}{n}\sum_{i=1}^{n} s_{i,j}} \qquad (1)$$

Where:

- $i = \{1, .., n\}$ represents the index of the sample that will be binned in the look-back window;
- $j = \{1, .., k\}$ represents the index of the metric, and k is the number of metrics that are being monitored;
- $s_{i,j}$ is the i-th sample value of the window of the *j-th* metric;
- $s_{i,j}^{'}$ is the i-th normalized sample value of the window of the *j-th* metric.

Data Binning Phase: it takes all normalized values and inserts them into a bin, which represents an interval for the data. The equation that represents the binning is:

$$b_{i,j} = \begin{cases} m, & s^{'}_{i,j} > r \\ \left\lfloor \dfrac{s^{'}_{i,j}}{r/m} \right\rfloor, & s^{'}_{i,j} \leqslant r \end{cases} \tag{2}$$

Where:

- $b_{i,j}$ is the corresponding bin index for $s^{'}_{i,j}$;
- r is the range of the normalized data. An $[0, r]$ interval is defined to represent the most representative data;
- m is the greatest bin index. Since the first index is 0, than there are *m+1* bins.

The bin index, $b_{i,j}$, of the normalized sample value $s^{'}_{i,j}$ will be the last bin index, m, if the sample value is greater than the range expected, r. The greatest sample values of the look-back window are placed into the m bin. The rest of the values are placed into the bins in the range $[0, r]$. To choose the adequate bin for those remaining values, we divide the normalized sample value by the ratio, which give us an idea of a fair placement of the values into m equal sized intervals. Finally, the bin index is the floor value of this division, as shown in (2).

Let C be the set of metrics being monitored. We may define $C = \{c_1, .., c_k\}, k = 1 \;..\; |C|$. We may monitor any number of metrics. For each metric being monitored, there is one sample value, and one corresponding bin. Examples of metrics at application-level are CPU and memory, and at network-level we may consider delay, bandwidth, and jitter, for instance.

Event Creation Phase: after normalizing and binning the data, the next step deals with representing the metrics as events. Those events are named *measurement events*, or *m-events*. One event, e_i, is a vector that contains the bins of all the *j* metrics analyzed. One m-event is defined as: $e_i = \langle b_{i,1}, b_{i,2}, ..., b_{i,k} \rangle$.

Entropy Computation and Aggregation Phase: we will compute the entropy of the events. Then, we may define *E* as the set containing all events in the current look-back window as:

$$E = \{e_1, e_2, ..., e_v\}$$

Where:

- Let v be the number of distinct events in the look-back window, then $v \neq n$ if there is more than one equal event;
- The event e_a is equal to event e_b if $b_{a,j} = b_{b,j}; \forall j \in [1, k], \forall b_{a,j} \in e_a, \forall b_{b,j} \in e_b, k = |e_a| = |e_b|$.

We will compute the entropy of E, $H(E)$. Firstly, we will count the number of occurrences of event e_i and represent it by $n_i, \forall i = [1, v]$. In the sequence, the local entropy may be calculated as stated in (3), where $^{n_i}/_n$ is the probability of occurring the event e_i [4].

$$H(E) = -\sum_{i=1}^{v} \frac{n_i}{n} log \frac{n_i}{n} \tag{3}$$

2) Processing of Entropy Time Series: The processing of the entropy time series consists of applying one or more methods striving to find out anomalous patterns. Those methods may be a combination of spike detection, signal processing, and subspace analysis, for instance.

B. Machine Learning Anomaly Detection

The unsupervised machine learning technique for anomaly detection technique is based on fitting the data to a Gaussian Distribution. The values with very low probability are considered anomalies. The goal of this probability analysis is to find out a probability threshold that maximizes the detection accuracy. In this section we will describe how one can implement such a technique.

We start by collecting measurements of the features (or *metrics*, in the context of network performance) that we call *training set* (TS). The next step is to fit a Gaussian distribution on the TS, calculating the probability of every value (the pair of features).

Given a *training set* $x^{(1)}, ..., x^{(m)}$ (where $x^{(i)} \in R_n$), let us estimate the Gaussian distribution for each of the features x_i. For each metric $i = 1, ..., n$, we need to find the parameters μ_i and σ_i^2 that fit the data in the i-th dimension $x_i^1, ..., x_i^m$ (i.e., the samples collected for metric i). The Gaussian distribution is given by (4), where μ is the mean and σ^2 is the variance. We estimate the parameters μ_i and σ_i^2 of the i-th metric by using (5) and (6), respectively [12].

$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

$$\mu_i = \frac{1}{m} \sum_{j=1}^{m} x_i^{(j)} \tag{5}$$

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^{m} (x_i^{(j)} - \mu_i)^2 \tag{6}$$

After calculating the mean and variance, we fit the data to the Gaussian model. Then, we observe which values have a very high probability according to the Gaussian distribution, and which have a very low probability. The low probability samples are anomalies. We predict which metric samples are anomalies by defining a *threshold*. We choose the threshold, ϵ, that maximizes the accuracy on a *cross validation set* [12].

Let the cross validation set be $CV = \{(x_{cv}^{(1)}, y_{cv}^{(1)}), \ldots, (x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})})\}$, where the label $y = 1$ corresponds to an anomalous metric sample, and $y = 0$ corresponds to a normal sample. For each cross validation element, we computed $p(x_{cv}^i)$, which is the mass probability of that element according to the Gaussian distribution. Let the vector of all of these probabilities be $P = \left\langle p(x_{cv}^{(1)}), \ldots, p(x_{cv}^{(m_{cv})}) \right\rangle$.

We define the threshold probability, ϵ, by selecting it at a range from the minimum and maximum values for $p(x_{cv}^i) \in P$. The ϵ value that maximizes the accuracy will be chosen as an anomaly indicator to the detection process [12].

IV. Efficient Entropy-based and Machine Learning-based Anomaly Detection

We argue that the anomaly detection based on the Gaussian model may help to set up the input parameters of the EbAT. We may use the EbAT to label the cross validation set adaptively. On the other hand, as time passes by, both techniques will work in synergy.

It is difficult to configure the input parameters of the EbAT technique, and to validate the identified alarms. On the other hand, from the machine learning perspective, it is difficult to label the samples without prior knowledge.

The proposed technique has 5 phases: (i) traffic capture; (ii) threshold estimation with machine learning; (iii) cloud services' monitoring; (iv) entropy estimation; (v) anomaly detection. We have summarized the processes of both techniques, described in Sections III-A and III-B, and how they interact in those phases in Figure 1. The idea is that the alarms generated by the EbAT will feed the subprocess of labeling the anomalous traffic packets, and that the threshold obtained by the machine learning-based technique will also become a proof when testing if the service metrics contains or not anomalous packets.

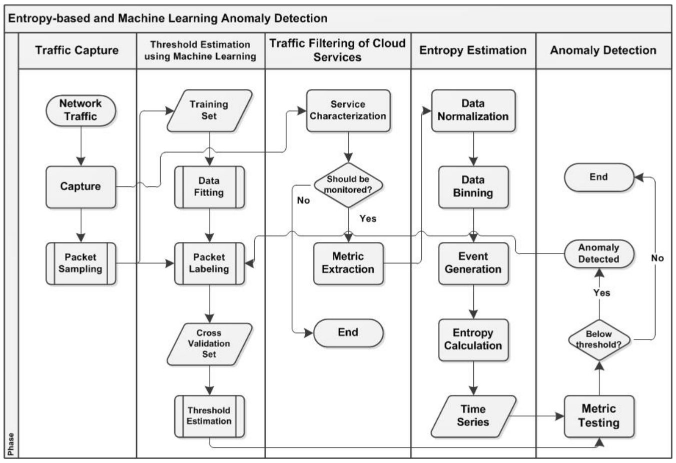


Figure 1. Proposed anomaly detection method.

V. Validation

A. Methodology

We are going to validate the accuracy of the detection system using the *F-measure* metric, also named *accuracy*, or *F1-score*. It represents the harmonic mean of the *precision* and *recall* metrics [13]. Those metrics are described along this section.

The **precision** metric is the ratio of the anomalies correctly detected and the total number of anomalies detected, either correct or wrong, shown in (7). Then, it characterizes the percentage of correctly detected anomalies; i.e., if a prediction algorithm has precision of 90%, then we understand that 90% of alerts are correct, and, thus, 10% of them are false positives, or *false alarm rate* (FAR, *1-Precision*).

$$Precision = \frac{\#\ of\ successful\ detections}{\#\ of\ total\ alarms} \qquad (7)$$

The **recall** metric, in turn, represents the ratio of anomalies correctly detected and the actual number of anomalies, as shown in (8). For example, if the recall metric is 55%, it denotes that 55% of the abnormalities were detected; consequently, there were 45% of missing alerts.

$$Recall = \frac{\#\ of\ successful\ detections}{\#\ of\ total\ anomalies} \qquad (8)$$

The analysis of those two metrics better expresses the degree of quality of the anomaly detector. How can we interpret the precision and recall metrics? When the same weights are assigned to the two metrics, one obtains the value of **F-measure** by (9). The larger the value of F-measure, the higher the quality predictor.

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (9)$$

B. Experiment Description

To analyze the behavior of the system, we adopted the model 2^3-factorial experimental design, which makes up a total of 8 different treatments. For each treatment, we have analyzed 3 factors (n, m, and r), each one varying at two levels, according to Table I. We have analyzed how the three factors and their interactions influenced the overall accuracy.

TABLE I. ANALYSED TREATMENTS.

Parameter	# Scenario							
	1	2	3	4	5	6	7	8
n	5	5	5	5	10	10	10	10
m	6	6	7	7	6	6	7	7
r	5	10	5	10	5	10	5	10

1) DoS Attack to Cloud Computing Services: We performed a VM-to-VM attack using the open-source tool Hping3 [14] within a cloud system running one application called Nutch, which digs up the Web searching for pages. The tool Hping3 allows us to generate arbitrary packets to flood a victim host. We set Hping3 on three VMs in a cloud to generate TCP SYN packets of the Hadoop application targeted to attack Hadoop ports on two victim VMs. The VMs are part of the same cloud, including one master node and one slave node, as shown in Figure 2. The timestamp of the attacks are detailed in Table II [15].

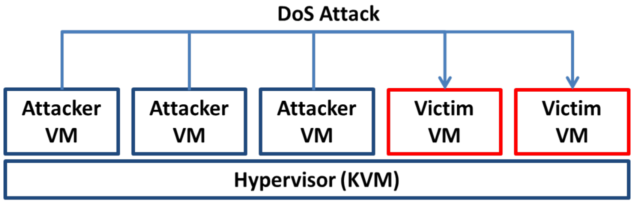


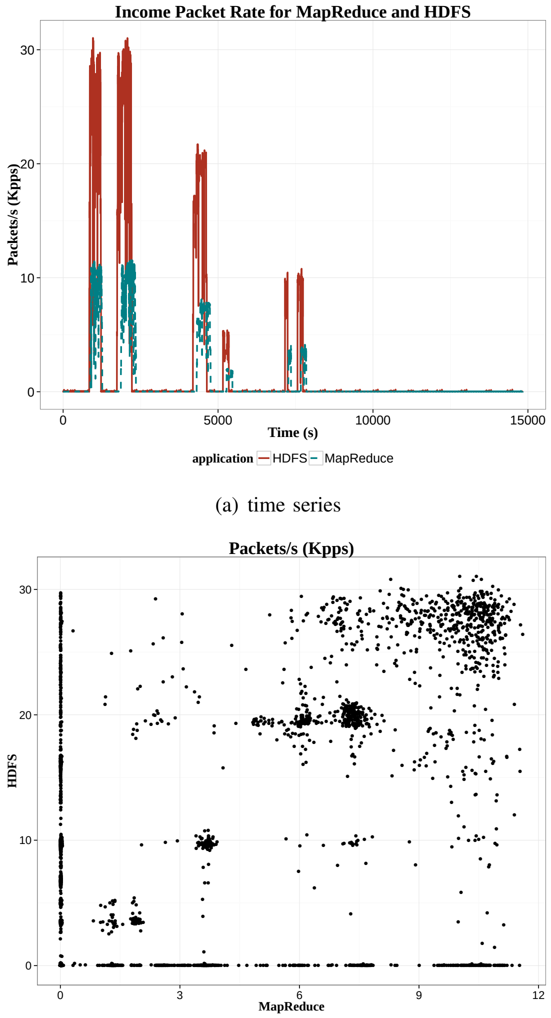
Figure 2. Nodes that are part of the DoS attack [15].

Then, we have measured the following metrics: (i) number of packets generated by the Hadoop Distributed File System (HDFS); (ii) number of packets generated by the MapReduce

TABLE II. TCP SYN FLOOD ATTACK LIST [15].

	Start Time	Finish Time	Attackers	Victims
1	18:23:32	18:29:10	VM3, VM4, VM5	VM1, VM2
2	18:40:14	18:47:58	VM3, VM4, VM5	VM1, VM2
3	19:20:35	19:28:15	VM4, VM5	VM1
4	19:36:29	19:39:35	VM3	VM1, VM2
5	20:09:37	20:11:06	VM4	VM1
6	20:16:36	20:19:08	VM4	VM1

application. The packet rate of those two applications is shown in Figure 3(a), and the scatterplot of the Map Reduce versus HDFS application packets is shown in Figure 3(b). We may observe that the packet rate series shown have correlation, however they are not synchronized.



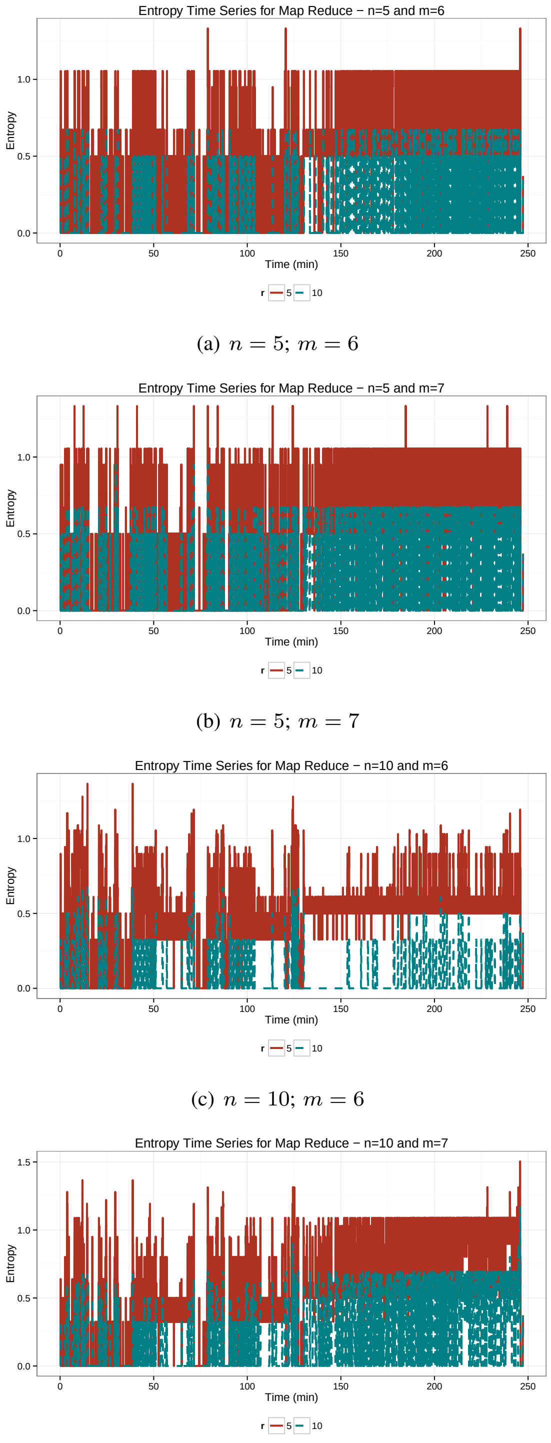
(a) time series

(b) scatterplot

Figure 3. MapReduce and HDFS packet rate.

We have calculated the entropy values for the pair of metrics measured in Figure 4. We may observe that the results are visually difficult to interpret, and establish the right parameters to propose the anomalous values.

Table III summarizes the results presented in Figure 4. We observe that the scenario 3 has the best results, i.e., the precision is 100 %. Although, there are others that provide good accuracy results as well, e.g., over 90 %.



(a) $n = 5$; $m = 6$

(b) $n = 5$; $m = 7$

(c) $n = 10$; $m = 6$

(d) $n = 10$; $m = 7$

Figure 4. Entropy time series with: $n = 5$ a, b); $n = 10$ (c, d).

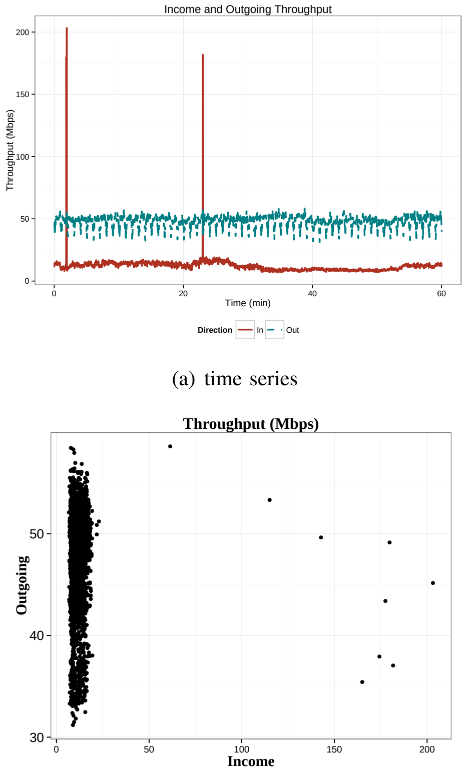
2) DoS Attack to a Network Provider Backbone: We have validated our work using a trace from the *Pohang University*

TABLE III. SUMMARY OF THE RESULTS OF THE ANOMALY DETECTION METRICS OF THE CLOUD DOS.

#Scenario	TP	FP	FN	Recall	Precision	F1
1	9967	711	1077	0.902481	0.933414	0.917687
2	543	0	10501	0.0491670	1	0.093726
3	**11044**	**0**	**0**	**1**	**1**	**1**
4	2181	0	8863	0.1974828	1	0.329830
5	10397	1749	647	0.941416	0.856002	0.896680
6	153	9	10891	0.013854	0.944444	0.027308
7	10638	1773	406	0.963238	0.857143	0.907099
8	2402	155	8642	0.217494	0.939382	0.353209

of Science and Technology (POSTECH) that contains traffic of a famous DDoS attack to government and commercial websites in South Korea in July 7th, 2009. Those attacks were probably launched by a special cyber warfare unit belonging to North Korean Army. During the attack, many computers in POSTECHs network campus were zombies. We have analysed one hour of network capture that contains the packets from the attack [16].

We measured and analysed two features of the network traffic: (i) the income throughput, and (ii) the outgoing throughput. The throughput series are depicted in Figure 5(a), and the scatterplot is in Figure 5(b). For this trace, we have applied the Gaussian model on a cross validation set to identify the threshold probability of having an anomalous sample. Then, we have used this value to predict which packets took part of the DoS attack.



(a) time series

(b) scatterplot

Figure 5. Income and outgoing throughput.

We have summarized the mean results obtained for the 8 different treatments in Table IV. We may realize that there are two treatments that contributes to the best accuracy results, which are the first and third scenario (in bold font). In this sense, we restricted our scope, and may also choose setup parameters from one of them both that lead this particular system to provide the best predictions about the presence of anomalies in the traffic. In our case, we selected the third scenario, which parameters are $n = 5$, $m = 7$, and $r = 5$. At those configurations, the F-measure reached 100 %.

TABLE IV. SUMMARY OF THE RESULTS OF THE ANOMALY DETECTION METRICS OF THE NETWORK PROVIDER.

#Scenario	TP	FP	FN	Recall	Precision	F1
1	**13**	**0**	**0**	**1**	**1**	**1**
2	0	0	13	0	NA	NA
3	**13**	**0**	**0**	**1**	**1**	**1**
4	0	0	13	0	NA	NA
5	12	9	1	0.923077	0.571429	0.705882
6	5	2	8	0.384615	0.714286	0.5
7	10	8	3	0.769231	0.555556	0.645161
8	0	0	13	0	NA	NA

We found out that the accuracy of the entropy-based anomaly detection mechanism itself directly depends on a further analysis of the traffic. Those assumptions, however, are too strong for a detection system and constitute barriers to the implementation of such a technique.

VI. FINAL REMARKS

Cloud computing systems have peculiar characteristics, such as aggregation of many different services, which makes it difficult to classify applications either by techniques based on packet payload signature matching, or probabilistic methods for the identification of traffic patterns, and profile of traffic normal behaviors. Those characteristics bring up challenges regarding online traffic monitoring and analysis, which should be done at wire speed while digging a high volume of data.

Choosing one optimal traffic anomaly detection technique is a complex task, because in order to have good results, we may have to know several characteristics of the traffic that are not known in practice. Another challenge is to develop high performance network packet sampling, and speeding up data processing, since the volume of traffic that traverses the cloud service provider is of the order of gigabits per second.

The obtained results show that when applying the EbAT itself, there is still the need to better analyze and comprehend the traffic patterns, finding out the normal behavior of the monitored systems, based on predictions using historical data, or feedback from experts on the business and network traffic, or by making new assumptions regarding the traffic. In this sense, we argue that the accuracy results may be improved by the aid of probability models, such as anomaly detection using the Gaussian model.

As a conclusion, we found out that it was still necessary to investigate new solutions to the cloud computing network anomaly detection problem. As future work, we intend to improve the anomaly detection mechanism by developing and analyzing new techniques to increase the detection accuracy, and to propose a parallel model for capturing and processing the network packets at wire speed.

References

[1] A. C. Oliveira, H. Chagas, M. Spohn, R. Gomes, and B. J. Duarte, “Efficient network service level agreement monitoring for cloud computing systems (to appear),” in Computers and Communications (ISCC), 2014 IEEE Symposium on, June 2014.

[2] S. Shetty, “Auditing and analysis of network traffic in cloud environment,” in Proceedings of the 2013 IEEE Ninth World Congress on Services, ser. SERVICES ’13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 260–267. [Online]. Available: http://dx.doi.org/10.1109/SERVICES.2013.42

[3] Q. Zhang, L. Cheng, and R. Boutaba, “Cloud computing: state-of-the-art and research challenges,” Journal of Internet Services and Applications, vol. 1, no. 1, Apr. 2010, pp. 7–18. [Online]. Available: http://www.springerlink.com/index/10.1007/s13174-010-0007-6

[4] C. Wang, “Ebat: online methods for detecting utility cloud anomalies,” in Proceedings of the 6th Middleware Doctoral Symposium, ser. MDS ’09. New York, NY, USA: ACM, 2009, pp. 4:1–4:6. [Online]. Available: http://doi.acm.org/10.1145/1659753.1659757

[5] C. Wang, V. Talwar, K. Schwan, and P. Ranganathan, “Online detection of utility cloud anomalies using metric distributions,” in 2010 IEEE Network Operations and Management Symposium - NOMS 2010. Ieee, 2010, pp. 96–103. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5488443

[6] C. Wang, K. Viswanathan, L. Choudur, V. Talwar, W. Satterfield, and K. Schwan, “Statistical techniques for online anomaly detection in data centers,” in Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on, May 2011, pp. 385–392.

[7] S. C. Wang, K. Q. Yan, and S. S. Wang, “Achieving High Efficient Agreement with Malicious Faulty Nodes on a Cloud Computing Environment,” Industrial Engineering, 2009, pp. 3–8.

[8] L. Benetazzo, G. Giorgi, and C. Narduzzi, “On the analysis of communication and computer networks by traffic flow measurements,” Instrumentation and Measurement, IEEE Transactions on, vol. 56, no. 4, Aug 2007, pp. 1157–1164.

[9] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, “An empirical evaluation of entropy-based traffic anomaly detection,” in Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, ser. IMC ’08. New York, NY, USA: ACM, 2008, pp. 151–156. [Online]. Available: http://doi.acm.org/10.1145/1452520.1452539

[10] Q. Quan, C. Hong-Yi, and Z. Rui, “Entropy based method for network anomaly detection,” in Dependable Computing, 2009. PRDC ’09. 15th IEEE Pacific Rim International Symposium on, Nov 2009, pp. 189–191.

[11] D. Smith, Q. Guan, and S. Fu, “An Anomaly Detection Framework for Autonomic Management of Compute Cloud Systems,” 2010 IEEE 34th Annual Computer Software and Applications Conference Workshops, Jul. 2010, pp. 376–381. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5615245

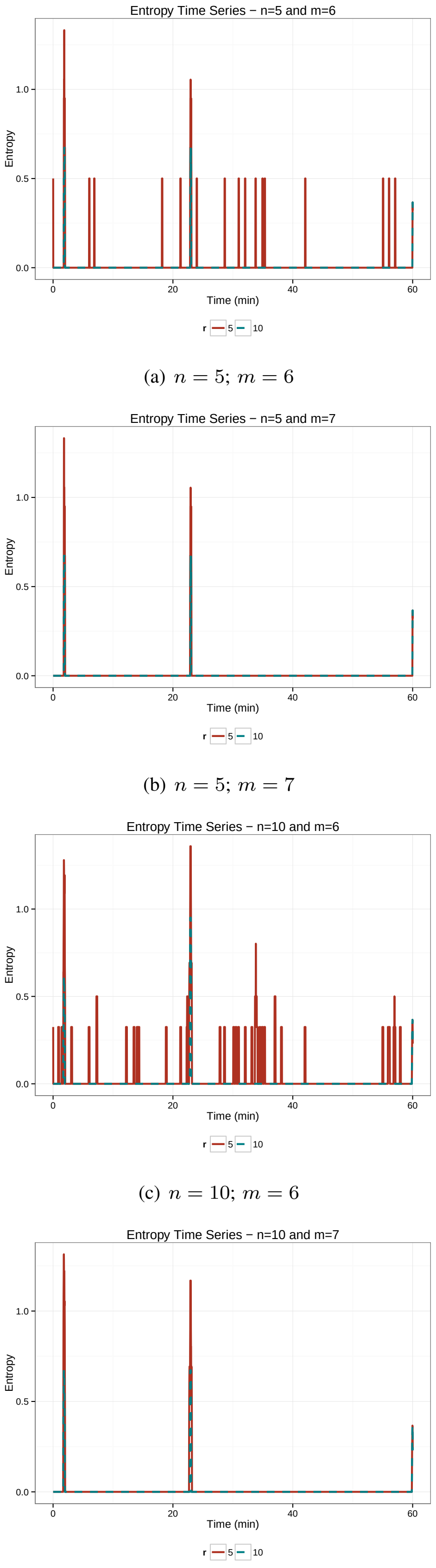
[12] A. Ng, “Machine learning: Anomaly detection,” Lecture Notes, Coursera Course, Standford University, September 2014. [Online]. Available: https://www.coursera.org/course/ml

[13] F. Salfner, M. Lenk, and M. Malek, “A survey of online failure prediction methods,” ACM Computing Surveys, vol. 42, no. 3, Mar. 2010, pp. 1–42. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1670679.1670680

[14] Hping, “Hping 3,” September 2014. [Online]. Available: http://www.hping.org/hping3.html

[15] D. L. Quoc, L. Yazdanov, and C. Fetzer, “Dolen: User-side multi-cloud application monitoring,” in Future Internet of Things and Cloud. IEEE, 2014.

[16] D. L. Quoc, T. Jeong, H. E. Roman, and J. W.-K. Hong, “Traffic dispersion graph based anomaly detection,” in Proceedings of the Second Symposium on Information and Communication Technology, ser. SoICT ’11. New York, NY, USA: ACM, 2011, pp. 36–41. [Online]. Available: http://doi.acm.org/10.1145/2069216.2069227



(a) $n = 5$; $m = 6$

(b) $n = 5$; $m = 7$

(c) $n = 10$; $m = 6$

(d) $n = 10$; $m = 7$

Figure 6. Entropy time series with: $n = 5$ (a, b); $n = 10$ (c, d).