# INTERNET 2014

The Sixth International Conference on Evolving Internet

ISBN: 978-1-61208-349-0

June 22 - 26, 2014

Seville, Spain

**INTERNET 2014 Editors**

Eugen Borcoci, University “Politehnica” Bucharest, Romania

Dirceu Cavendish, Kyushu Institute of Technology, Japan

# INTERNET 2014

# Foreword

The Sixth International Conference on Evolving Internet (INTERNET 2014), held between June 22-26, 2014 - Seville, Spain, dealt with challenges raised by evolving Internet making use of the progress in different advanced mechanisms and theoretical foundations. The gap analysis aimed at mechanisms and features concerning the Internet itself, as well as special applications for software defined radio networks, wireless networks, sensor networks, or Internet data streaming and mining.

Originally designed in the spirit of interchange between scientists, the Internet reached a status where large-scale technical limitations impose rethinking its fundamentals. This refers to design aspects (flexibility, scalability, etc.), technical aspects (networking, routing, traffic, address limitation, etc), as well as economics (new business models, cost sharing, ownership, etc.). Evolving Internet poses architectural, design, and deployment challenges in terms of performance prediction, monitoring and control, admission control, extendibility, stability, resilience, delay-tolerance, and interworking with the existing infrastructures or with specialized networks.

We take here the opportunity to warmly thank all the members of the INTERNET 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 who dedicated much of their time and efforts to contribute to INTERNET 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 are grateful to the members of the INTERNET 2014 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that INTERNET 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of the evolving internet.

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

**INTERNET 2014 Chairs:**

**INTERNET Advisory Committee**

Eugen Borcoci, University “Politehnica” Bucharest, Romania
Abdulrahman Yarali, Murray State University, USA
Mark Yampolskiy, Vanderbilt University, USA
Vladimir Zaborovsky, Technical University - Saint-Petersburg, Russia
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Evangelos Kranakis, Carleton University, Canada

Danny Krizanc, Wesleyan University-Middletown, USA
Natalija Vlajic, York University - Toronto, Canada
Krzysztof Walkowiak, Wroclaw University of Technology, Poland
Junzo Watada, Waseda University - Fukuoka, Japan
Robert van der Mei, Centrum Wiskunde & Informatica, The Netherlands

**INTERNET Industrial/Research Chairs**

Jerome Galtier, Orange Labs, France
Martin Dobler, FH VORARLBERG - Dornbirn, Austria
Abdelmajid Khelil, Huawei Research, Germany
Tingyao Wu, Alcatel-Lucent/Bell Labs, USA

# INTERNET 2014

## Committee

### INTERNET Advisory Committee

Eugen Borcoci, University “Politehnica” Bucharest, Romania
Abdulrahman Yarali, Murray State University, USA
Mark Yampolskiy, Vanderbilt University, USA
Vladimir Zaborovsky, Technical University - Saint-Petersburg, Russia
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Evangelos Kranakis, Carleton University, Canada
Danny Krizanc, Wesleyan University-Middletown, USA
Natalija Vlajic, York University - Toronto, Canada
Krzysztof Walkowiak, Wroclaw University of Technology, Poland
Junzo Watada, Waseda University - Fukuoka, Japan
Robert van der Mei, Centrum Wiskunde & Informatica, The Netherlands

### INTERNET Industrial/Research Chairs

Jerome Galtier, Orange Labs, France
Martin Dobler, FH VORARLBERG - Dornbirn, Austria
Abdelmajid Khelil, Huawei Research, Germany
Tingyao Wu, Alcatel-Lucent/Bell Labs, USA

### INTERNET 2014 Technical Program Committee

Jemal Abawajy, Deakin University - Victoria, Australia
Cristina Alcaraz, University of Malaga, Spain
Onur Alparslan, Osaka University, Japan
Mercedes Amor, University of Malaga, Spain
Olivier Audouin, Alcatel-Lucent Bell Labs, France
Jacques Bahi, University of Franche-Comté, France
Nik Bessis, University of Derby, UK
Maumita Bhattacharya, Charles Sturt University - Albury, Australia
Bruno Bogaz Zarpelão, State University of Londrina (UEL), Brazil
Eugen Borcoci, University “Politehnica” Bucharest, Romania
Christian Callegari, University of Pisa, Italy
Maya Carrillo Ruiz, Benemérita Universidad Autónoma de Puebla (BUAP), Mexico
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Antonio Celesti, University of Messina, Italy
Yue-Shan Chang, National Taipei University, Taiwan

Claude Chaudet, Telecom ParisTech, France
Shiping Chen, Sybase Inc., USA
Weifeng Chen, California University of Pennsylvania, USA
Yaw-Chung Chen, National Chiao Tung University, Taiwan
Young-Long Chen, National Taichung University of Science and Technology, Taiwan
Albert M. K. Cheng, Member, University of Houston, USA
Hongmei Chi, Florida A&M University, USA
Been-Chian Chien, National University of Tainan, Taiwan
Andrzej Chydzinski, Silesian University of Technology - Gliwice, Poland
José Alfredo F. Costa, Federal University, UFRN, Brazil
Jean-François Couchot, Universitéde Franche-Comté (LIFC), France
Guillermo Diaz-Delgado, Universidad Autónoma de Querétaro (UAQ) / Queretaro State University (UAQ), Mexico
Ioanna Dionysiou, University of Nicosia, Cyprus
Phan-Thuan Do, Hanoi University of Science and Technology, Vietnam
Martin Dobler, FH VORARLBERG - Dornbirn, Austria
Yingfei Dong, University of Hawaii, USA
Mohamed Dafir El Kettani, ENSIAS - Université Mohammed V-Souissi - Rabat, Morocco
Marc Fabri, Leeds Metropolitan University, UK
Zongming Fei, University of Kentucky, USA
Giancarlo Fortino, University of Calabria - Rende, Italy
Steffen Fries, Siemens AG, Germany
Song Fu, University of North Texas - Denton, USA
Marco Furini, Universita' di Modena e Reggio Emilia, Italy
Jerome Galtier, Orange Labs, France
Filippo Gandino, Politecnico di Torino, Italy
Miguel Garcia, Polytechnic University of Valencia, Spain
Bezalel Gavish, Southern Methodist University - Dallas, USA
S.K. Ghosh, Indian Institute of Technology - Kharagpur, India
Victor Govindaswamy, Texas A&M University-Texarkana, USA
Annie Gravey, Technopôle Brest Iroise, France
Javier Gutierrez, University of Seville, Spain
Frederic Guyard, Orange Labs, France, France
Frans Henskens, University of Newcastle, Australia
Ching-Hsien Hsu, Chung Hua University, Taiwan
Wladyslaw Homenda, Warsaw University of Technology, Poland
Pao-Ann Hsiung, National Chung Cheng University, Taiwan
Ching-Hsien Hsu, Chung Hua University, Taiwan
Fu-Hau Hsu, National Central University, Taiwan
Yongjian Hu, University of Warwick, UK
Yo-Ping Huang, National Taipei University of Technology - Taipei, Taiwan
Marc Jansen, University of Applied Sciences Ruhr West, Germany
Terje Jensen, Telenor Corporate Development - Fornebu / Norwegian University of Science and Technology - Trondheim, Norway

Young-Sik Jeong, Dongguk University Seoul, Korea
Epaminondas Kapetanios, The University of Westminster, UK
Abdelmajid Khelil, Huawei Research, Germany
Muhammad Khurram Khan, King Saud University, Saudi Arabia
Donghyun (David) Kim, North Carolina Central University, USA
Wojciech Kmiecik, Wroclaw University of Technology, Poland
Ren-Song Ko, National Chung Cheng University, Taiwan
Igor Kotenko, SPIIRAS, Russia
Vitomir Kovanovic, Simon Fraser University - Surrey, Canada
Constantine Kotropoulos, Aristotle University of Thessaloniki, Greece
Evangelos Kranakis, Carleton University, Canada
Danny Krizanc, Wesleyan University-Middletown, USA
Michal Kucharzak, Wroclaw University of Technology, Poland
KP Lam, University of Keele, UK
Clement Leung, Hong Kong Baptist University, Hong Kong
Juan Li, North Dakota State University, USA
Fidel Liberal Malaina, University of the Basque Country, Spain
Xingcheng Liu (刘星成), Sun Yat-sen University - Guangzhou, China
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Seng Loke, La Trobe University, Australia
Isaí Michel Lombera, University of California - Santa Barbara, USA
Juan M. Lopez-Soler, University of Granada, Spain
Henrique João Lopes Domingos, New University of Lisbon & CITI Research Center, FCT/UNL - Lisbon, Portugal
Damien Magoni, University of Bordeaux - Talence, France
Gregorio Martinez, University of Murcia, Spain
Albena Mihovska, Aalborg University, Denmark
Sangman Moh, Chosun University - Gwangju, South Korea
Paul Mueller, University Kaiserslautern, Germany
Ethiopia Nigussie, University of Turku, Finland
Samuel Nowakovski, LORIA, France
Masaya Okada, Shizuoka University, Japan
Luis M. Oliveira, Instituto de Telecomunicações, Portugal
Pasquale Pace, University of Calabria, Italy
Janne Parkkila, Lappeenranta University of Technology, Finland
Luigi Patrono, University of Salento, Italy
Ilia Petrov, Reutlingen University, Germany
Angel P. del Pobil, Jaume I University, Spain
Marek Reformat, University of Alberta - Edmonton, Canada
Hamed Sadeghi Neshat, University of British Columbia
Abdel-Badeeh M. Salem, Ain Shams University Abbasia - Cairo, Egypt
Paul Sant, University of Bedfordshire, UK
José Santa, University of Murcia, Spain
Peter Schartner, University of Klagenfurt, Austria

Roman Y. Shtykh, CyberAgent, Inc., Japan
Ramesh Sitaraman, University of Massachusetts - Amherst,USA
Dimitrios Serpanosm ISI/R.C. Athena & University of Patras, Greece
Yang Song, IBM Research, USA
Pedro Sousa, University of Minho, Portugal
Neuman Souza, Federal University of Ceara, Brazil
Shensheng Tang, Missouri Western State University, USA
Juan E. Tapiador, Universidad Carlos III de Madrid, Spain
Ruppa K. Thulasiram, University of Manitoba - Winnipeg, Canada
Parimala Thulasiraman, University of Manitoba - Winnipeg, Canada
Herwig Unger, FernUniversitaet in Hagen, Germany
Muhammad Usman, Auckland University of Technology, New Zealand
Robert van der Mei, Centrum Wiskunde & Informatica, The Netherland
Massimo Villari, University of Messina, Italy
Natalija Vlajic, York University - Toronto, Canada
Krzysztof Walkowiak, Wroclaw University of Technology, Poland
Junzo Watada, Waseda University - Fukuoka, Japan
Sabine Wittevrongel, Ghent University, Belgium
Kui Wu, University of Victoria, Canada
Tingyao Wu, Alcatel-Lucent/Bell Labs, USA
Zhengping Wu, University of Bridgeport, USA
Mudasser F. Wyne, National University - San Diego, USA
Bin Xie, InfoBeyond Technology LLC - Louisville, USA
Mark Yampolskiy, Vanderbilt University, USA
Zhenglu Yang, The University of Tokyo, Japan
Chuan Yue, University of Colorado - Colorado Springs, USA
Habib Zaidi, Geneva University Hospital, Switzerland
Zhao Zhang, Iowa State University, USA
Weiying Zhu, Metropolitan State University of Denver, USA
Cliff C. Zou, University of Central Florida - Orlando, USA

**ACCESS 2014**

**ACCESS Advisory Committee**

Alessandro Bogliolo, Università di Urbino, Italy
Mark Perry, University of New England in Armidale, Australia
Abdulrahman Yarali, Murray State University, USA
Ljiljana Trajkovic, Simon Fraser University - Burnaby, Canada
Ronit Nossenson, Jerusalem College of Technology, Israel

## ACCESS Industrial/Research Chairs

Deepak Kataria, IPJunction Inc. - Bridgewater, USA
Michael Bahr, Siemens AG - Munich, Germany
George Korinthios, COSMOTE - Mobile Telecommunications S.A. - Athens, Greece
Yunxin (Jeff) Li, IP Australia, Australia
Serena Elisa Ponta, SAP Lab - Mougins, France
Alejandro Cordero, Amaranto Consultores, Spain
Arief Mustain, PT Telekomunikasi Indonesia, Tbk, Indonesia

## ACCESS Publicity Chairs

Andrzej Beben, Warsaw University of Technology, Poland
Kwang Ok Kim, ETRI, Korea

## ACCESS 2014 Technical Program Committee

Hojjat Adeli, The Ohio State University, USA
Konstantin Avratchenkov, INRIA, France
Michael Bahr, Siemens AG - Munich, Germany
Andrzej Beben, Warsaw University of Technology, Poland
Alessandro Bogliolo, Università di Urbino, Italy
Fernando Boronat Seguí, Polytechnic University of Valencia - Gandia, Spain
Alejandro Cordero, Amaranto Consultores, Spain
Istvan Frigyes, Budapest University of Technology and Economics, Hungary
Emiliano Garcia-Palacios, Queens University - Belfast, UK
Antonis M. Hadjiantonis, University of Cyprus, Cyprus
Tauseef Jamal, University Lusofona - Lisbon, Portugal
Georgios Karagiannis, University of Twente, The Netherland
Deepak Kataria, IPJunction Inc. - Bridgewater, USA
Sokratis K. Katsikas, University of Piraeus, Greece
Gul Muhammad Khan, University of Engineering and Technology Peshawar, Pakistan
George Korinthios, COSMOTE - Mobile Telecommunications S.A. - Athens, Greece
Pandelis Kourtessis, University of Hertfordshire, U.K.
Trung-Thanh Le, Hanoi University of Natural Resources and Environment, Vietnam
Gyu Myoung Lee, Institut Telecom / Telecom SudParis, France
Yunxin (Jeff) Li, IP Australia, Australia
Enjie Liu, University of Bedfordshire, UK
Olaf Maennel, Loughborough University, UK
Amin Malekmohammadi, University of Nottingham Malaysia Campus, Malaysia
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Elsa María Macías López, University of Las Palmas de Gran Canaria, Spain
Jon Matias, University of the Basque Country (UPV/EHU), Spain
Armando Nolasco Pinto, Instituto de Telecomunicações / Universidade de Aveiro, Portugal

## Copyright Information

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

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

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

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

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

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

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

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

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

## Table of Contents

# A New Approach to Anomaly Detection based on Possibility Distributions

Joseph Ndong
Department of Mathematics and Computer Science,
University Cheikh Anta Diop of Dakar, Sénégal
Email: joseph.ndong@ucad.edu.sn

***Abstract*—This paper presents a new approach for anomaly detection based on possibility theory for normal behavioral modeling. Combining subspace identification algorithms and Kalman filtering techniques could be a good basis to find a suitable model to build a decision variable where, a new decision process can be applied to identify anomalous events. A robust final decision scheme can be built, by means of possibility distributions to find the abnormal space where anomalies happen. Our system uses a calibrated state space dynamical linear model where the model's parameters are found by the principal component analysis framework. The multidimensional Kalman innovation process is used to build the unidimensional decision variable. Thereafter this variable is clustered and possibility distributions are used to separate the clusters into normal and abnormal spaces when anomalies happen. We had studied the false alarm rate *vs.* detection rate trade-off by means of the Receiver Operating Characteristic curve to show the high performance obtained via this new methodology against other approaches. We validate the approach over different realistic network traffic.**

***Index Terms*—Anomaly detection, GMM, probability-possibility theory, subspace identification, PCA, Kalman filter.**

## I. INTRODUCTION

Kalman filter based techniques first calibrate a Maximum-Likelihood based model for normal behavior modeling for the entropy reduction step [10][11][12]. Thereafter the decision variable is obtained as the filter innovation process. Analyzing residual for anomaly detection can be a good approach, since in favorable conditions, this process is assumed to be a zero mean gaussian white noise. However, if we believe that anomalies can cause low, high or abrupt changes in the traffic, this can attempt to appear in different statistical properties in the residual, making us to believe that, this signal is instead an ensemble of normal distributions. So, it will be interesting to take into account the residual process and, try to build a few set of (normal/abnormal) clusters. Finally, our attention can be put on the abnormal clusters to track anomalies.

Principal component analysis (PCA) approach [7][8][15] provides very good model of normal behavior with strong differentiation with abnormal behavior. However it is weaken by its high sensitivity to non-stationarity and parameter settings. Whereas Kalman filtering approach is inherently more robust to some level of non stationarity in the data because of its feedback structure. However, the main weakness in the approach proposed initially in [10] is within the Maximum Likelihood estimation that fails in capturing the essential properties of the normal behavior. The previous analysis lead us to believe that combining a PCA based normal model with Kalman filtering step, can be a good basis for building a suitable decision variable where possibilistic test could be applied for anomaly detection. In this work, we show that *subspace identification* algorithm can be used in combination of a Kalman filter to build the decision variable.

In this work, we are interested in anomaly detection based robust unsupervised clustering. If we assumed that, generally, anomalies might be rare, one can build a *few* number of clusters and try to find them in some of these classes. There are two major informations which seem to be relevant for detecting true anomalous events, and which we want to exploit here, to build a robust anomaly detector. One can determine clearly the posterior probability of a data sample being distributed in the different clusters, but we have no idea of the probability of generating the clusters themselves. Thus, using possibility distribution to estimate the degree a cluster can be seen as ”possible”, should be a great interest for anomaly detection. Thus we follow [4] to characterize the unknown probabilities of generating a set of clusters by *simultaneous confidence intervals* with a given confidence level $1-\alpha$. Thereafter these intervals will be used to calculate possibility distributions (degree of possibility) for each cluster. This operation will have the ability to separate the different classes into normal and abnormal sub spaces. It will be, at the same time, necessary to have at hand the possibility distributions for the data sample to recover a critical value of the cluster possibility degree (which we will use to determine the normal and abnormal clusters).

The organization of this paper is as follows. Section III deals with the methodology we adopt in our anomaly detection scheme. In Section IV, we validate our approach by showing efficient results. Section V concludes the work and fix some ideas for future study.

## II. RELATED WORKS

In our knowledge, this work presents the first approach that deals with possibility theory to build an anomaly detector for communication networks. Generally, in the literature, the proposed approaches are based on Bayesian inference i.e., probabilistic solutions. We have developed recently some techniques for anomaly detection using statistical approaches. In [23], the proposed method to detect anomalous events is based on gaussian mixture model (GMM) for clustering. Thereafter, we proposed a hidden markov model (HMM) coupled with the Viterbi algorithm to subdivide the space into two other sub spaces: the first containing a few number of cluster data corresponding to the abnormal sub space and the second one containing the majority of the data and corresponding to the normal space. However in that study, there is a great challenge to calibrate properly the GMM since it is necessary to run the model several times to achieve model convergence. The same problem occurs when searching for the best parameters of the HMM in order to learn about spacial and temporal correlations between the GMM clusters, in order to classify them into two or more states. In [27] the monitoring system is also based on the coupling of a GMM and HMM and the same problems arise. The searching of the best number of the HMM parameters need a thorough calibration of several models and the choice of the best model is based on the transition matrix. One should have high probabilities in the main diagonal of this matrix to decide the model selection. However ”high probability” was not defined more suitably. Our present work deal with these problems by proposing a new scheme based on possibility theory to separate the GMM clusters into two sub-spaces corresponding to the normal and abnormal regions. Our model does not necessitate multiple re-calibrations and the methodology to build a threshold to separate the space into other sub spaces is more reliable. In this work, we show the advantage to use the framework of possibility theory which is more reliable to characterize the probability of generating the clusters themselves, since we do not know their real distributions. This operation have the great advantage to mark a cluster as normal or abnormal. This findings was not achieve in the previous studies.

## III. Normal behavior modeling

An anomaly detector is generally built using a normal behavior model. As network traffic is a dynamical signal, one would like to build a dynamical normal behavior model. A classical approach to model dynamical signal is using Linear Time Invariant State-Space (LTISS) [17] model, representing input-output multivariate data sequences, as shown in the following difference equations:

$$\begin{cases} x_{t+1}=Ax_t + Bu_t + w_t \\ y_t \ =Cx_t + Du_t + v_t \end{cases} \tag{1}$$

In ( 1), the system state $x_t$, the measurable output $y_t$ and the input $u_t$ are multi-dimensional vectors of appropriate dimensions. The noise processes are assumed to be uncorrelated zero-mean gaussian white-noise processes with covariance matrices $cov(w_t) = Q$ and $cov(v_t) = R$, respectively. The input signal and the process noise are assumed to be statistically independent.

### A. How to build the Decision Variable ?

Calibrating a normal behavior model need to finding the values $(A, B, C, D)$ and $(Q, R)$ that fit better a learning set containing signals gathered over a period where, no anomalies have happened. To calibrate these system quantities, we follow the methodology described in [16], where subspace identification algorithms are presented to be a valuable tool to identify the state space parameters. Sub space algorithms have the ability to provide accurate state space models for multivariate linear systems and to retrieve system related matrices as sub spaces of projected data matrices. This means that the Kalman filter states can be recovered from the given input-output data. The identification problem is essentially characterized by the extraction of these matrices from input-output data, by using *QR* factorization and Singular Value Decomposition (SVD). In this work, we use the sub space identification algorithms based on Multivariable Output-Error State Space (MOESP) approach. The main idea for models based on MOESP method is to reconstruct the *past* input-output and *future* input-output data. The multi-dimensional output response $Y$ and input $U$ are first transformed into block Hankel matrices. Then the MOESP algorithm performs the compression of a compound matrix using the input and output Hankel matrices, into a lower triangular matrix by means of orthogonal transformations and *QR* decomposition. Thereafter, the column space of specific sub matrices of the resulting lower triangular factor approximates the column space of the extended observability matrix in a convenient way. Thereafter PCA can be computed by means of SVD technique and solution of a set of linear equations can then be performed to find the deterministic components. There is many raisons to use sub space model identification methods (SMI) for state space parameter learning: i) when correctly implemented, SMI algorithms are fast, despite the fact that they use QR and SVD decomposition. They are faster than classical identification methods, such as Prediction Error Methods, because they are not iterative ii) numerical robustness is guaranteed precisely due to the well-understood algorithms obtained from numerical linear algebra iii) the user will never be confronted with problems such as: lack of (slow) convergence, numerical instability, local minima and sensitivity of initial estimates iv) the reduced model can be obtained directly from input-output data, without having to compute first the high order model, one is always inclined to obtain models with as low as an order as possible.

In subspace model identification approach, one key step is the approximation of a structural subspace from spaces defined by Hankel matrices, constructed from the input-output data. For the LTISS system, the matrix pair $\{A, C\}$ is assumed to be observable, which implies that all modes in the system can be observed in the output $y_t$ and can thus be identified; this also implies that the rank of the extended observability matrix is equal to $N$. The system $\{A, (BQ^{1/2})\}$ is assumed **to be controllable i.e., the modes** of the system $\{A, Q^{1/2}\}$ are assumed to be stable. That structured subspace is the extended observability matrix $\Gamma_i$ (where $i$ denotes the number of block rows), which is defined as:

$$\Gamma_i = \left[ \begin{array}{ccccc} C & CA & CA^2 & \dots & CA^{i-1} \end{array} \right]^T \tag{2}$$

From the LTISS, we can re-organize the data by the following algebraic relationships:

$$Y_{k,i,j} = \Gamma_i X_{k,j} + H_i U_{k,i,j} + T_i E_{k,i,j} \tag{3}$$

For simplicity, we can rewrite the above equation as:
$Y = \Gamma_i X + H_i U + T_i E$.
In ( 3), $Y_{k,i,j}$, $U_{k,i,j}$, and $E_{k,i,j}$ are block Hankel matrices with $i$ block rows and $j$ block columns of the form:

$$\mathbf{Y_{k,i,j}} = \begin{bmatrix} y_k & y_{k+1} & \cdots & y_{k+j-1} \\ y_{k+1} & y_{k+2} & \cdots & y_{k+j} \\ \vdots & \vdots & \vdots & \vdots \\ y_{k+i-1} & y_{k+i} & \cdots & y_{k+j+i-2} \end{bmatrix} \tag{4}$$

The user-defined subscript $i$ should be large enough, i.e larger than the order $N$ of the system. The Hankel matrices $U_{k,i,j}$, and $E_{k,i,j}$ are defined in the same way.
$H_i$ and $T_i$ are Toeplitz matrices defined as:

$$\mathbf{H_i} = \begin{bmatrix} D & 0 & \ldots & 0 \\ CB & D & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{i-2}B & \ldots & CB & D \end{bmatrix} \tag{5}$$

$$\mathbf{T_i} = \begin{bmatrix} I & 0 & \ldots & 0 \\ CK & I & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{i-2}K & \ldots & CK & I \end{bmatrix} \tag{6}$$

The state sequence matrix $X_{k,j}$ is defined as:

$$X_{k,j} = [x_k + x_{k+1} + x_{k+2} + \ldots + x_{k+j-1}] \tag{7}$$

In MOESP algorithm, one has to determine the *extended observability matrix* by means of *orthogonal projections* on sub spaces span by $U$ columns. Thereafter, one can extract the measurement matrix $C$ by using the first block from the extended observability matrix $\Gamma_i$, and the state matrix $A$ is obtained by using the following formulas:

$$\Gamma_{2:i+1} = \begin{bmatrix} CA & \ldots & CA^i \end{bmatrix}^T = \Gamma_i A \tag{8}$$

More precisely: $A = \Gamma_i^{\dagger}\Gamma_{2:i+1}$, where $(.)^{\dagger}$ denotes the Moore-Penrose pseudo inverse matrix.

To find the observability matrix, one has to first split the input-output data into distinct past and future input-output sequences, and thereafter build a lower triangular matrix by means of orthogonal transformations using the *QR* factorization as follows:

$$\begin{bmatrix} U_{1,i,j} \\ U_{i+1,i,j} \\ Y_{1,i,j} \\ Y_{i+1,i,j} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ R_{21} & R_{22} & 0 & 0 \\ R_{31} & R_{32} & R_{33} & 0 \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_1^T \\ Q_1^T \\ Q_1^T \end{bmatrix} \tag{9}$$

This *QR* factorization is only valid for the case where the input signal is a zero-mean white noise, as in our case. For an arbitrary input signal, see [17] for an appropriate *QR* factorization. Thereafter SVD is computed as:

$$[R_{42} \quad R_{43}] = \hat{Q}_s \hat{\Sigma}_s \hat{V}_s^T + \hat{Q}_N \hat{\Sigma}_N \hat{V}_N^T = \Gamma_i X \prod_{U^T}^{\perp} \tag{10}$$

where $X \prod_{U^T}^{\perp}$ has full rank N and $\prod_{U^T}^{\perp}$ denotes the orthogonal projections on the lines of the null space of U.

$\Gamma_i$ is estimated from $\hat{Q}_s$, which has the $N$ principal left singular vectors corresponding to the most significant singular values. After finding an approximation of the Toeplitz $H_\gamma$ matrix (see [17]), the matrices $B$ and $D$ are computed from least-square solution of the following over determined system [18]:

$$\begin{bmatrix} R_{31} & R_{42} \end{bmatrix} \cong H_\gamma \begin{bmatrix} R_{11} & R_{22} \end{bmatrix} \tag{11}$$

After finding the above matrices by calibrating a predictive model by means of PCA, the model described by equation 1 is re-used, and we perform Maximum Likelihood using a Kalman filter, in order to build the *decision variable* using the multi-dimensional innovation process obtained as output of the Kalman filter. The ***one-dimensional*** decision variable (DV) process is obtained by applying the formulas:

$$decisionvariable = e(t)^T V e(t) \tag{12}$$

where the matrix $V$ (obtained as output of the Kalman filter) is the inverse of the **variance** of the multi-dimensional innovation process $e(t)$, $T$ denotes the transpose.

The Maximum Likelihood framework can be built by running the *predictor-corrector* iterative algorithm using two steps: *prediction* comes in the *time update* phase, and *correction* in the *measurement update* phase. Due to lack of space, we do not put in the text the different equations related to these two steps of the Kalman filter. The reader can find the calibration in our previous works in [23][27] and in other studies [11][12][21].

### B. How to build to normal subspace ?

We aim in this paper to learn residuals (i.e.,, the innovation process as output of the Kalman filter) for anomaly detection. Generally, it is assumed that the Kalman residual is a zero mean white gaussian noise. But it is often false to consider this assertion as a whole property of this process. In place, we are assuming that the real distribution of the innovation process is a mixture of normal distributions. We can simply calibrate a gaussian mixture model (GMM) [25], to organize the data in few number of clusters (i.e gaussian components). Anomalies might then appear is some of these gaussian components, and if one can carefully extract the potentially "abnormal" clusters (the remaining being labelled as "normal"), a basic test should be applied to detect the anomalous events. We will see in this work that, this aim can be achieved via the use of possibility distributions. First, we will use the sophisticated High Dimensional Data Clustering (HDDC) method presented by Bouveyron and al. [26], which is robust to find the best number of clusters and model parameters with low complexity.

*1) Clustering operation:* Generally, measured observations in communication networks are high dimensional. A popular approach to perform unsupervised clustering is to use gaussian mixture model (GMM), which rely on the assumption that each class can be represented by a gaussian density. This method supposes that observations $\{x_1, \ldots, x_N\}$ are independent realizations of a random vector $X \in \mathbb{R}^p$ with density :

$$f(x,\theta) = \sum_{k=1}^{K} \pi_k \phi(x; \mu_k, \Sigma_k) \tag{13}$$

where $\pi_k$ denotes the mixture proportion of the $kth$ component and $\phi$ is the gaussian density parametrized by the mean $\mu_k$ and the covariance matrix $\Sigma_k$. The classical approach (the well-know quadratic discriminant analysis-QDA) requires the estimation of a very large number of parameter (proportional to $p^2$, the number of variables in the dataset), and therefore faces to numerical problems in high dimensional spaces. In addition, classical gaussian mixture models show a disappointing behavior when the size of the training dataset is too small compared to the number of parameters to estimate. To avoid overfitting, it is necessary to find a balance between the number of parameters to estimate and the generality of the model. The HDDC acts in this way and the approach assumes that high-dimensional data live around specific sub spaces with a dimension lower than $p$. Bouveyron and al. have introduced a new parametrization of the gaussian mixture model which takes into account the specific sub space around which each cluster is located and, therefore limits the number of parameters to estimate. Many kind of models are proposed and the estimation of the model's parameters is done via the Expectation-Maximization (EM) algorithm and, some variants as Classification EM (CEM) for faster convergence and Stochastic EM (SEM) to avoid initialization problem. The intrinsic dimension of each cluster is determined automatically with the scree test of Catell [20], where we search a break in the curse corresponding to a local maxima. The best number of clusters can be derived by means of the Bayesian information criterion (BIC) [24]. When running the HDDC for a given model $r$, some additional parameters are added that may increase the likelihood. This operation can cause overfitting which can be avoid by the BIC criterion which introduces a penalty term for the number of parameters in the model. With the set of estimated models, the one with the lower value of BIC is preferred.

After finding $K$ clusters for the multi-dimensional innovation process (kalman residual), we applying the result to the unidimensional decision variable built in (12), to put it into $K$ clusters. It is simple to achieve this, because the HDDC gives the cluster labels (as a sequence of $N$ mixing symbols $[1, 2, \ldots, K]$). At the same time, the HDDC clustering phase gives us a $n \times K$ matrix representing the posterior probabilities $t_{ik}$ that the observation $i$ belongs to the cluster $k$, which can be used to calculate the possibility distribution for the data sample as defined in (17).

*2) Possibility theory as a tool to build normal space:* Our aim here is to infer possibility distribution from data to build the normal space. Dubois and Prade have built a procedure [1][2][3] which produces the most specific possibility distribution among the ones dominating a given probability distribution. In this paper, this method is generalized to the case where the probabilities (of generating the clusters) are **unknown**. We assume the above clusters have been generated from an unknown probability distribution. It is proposed to characterize the probabilities of generating the different clusters by *simultaneous confidence intervals* with a given confidence level $1 - \alpha$. A procedure for constructing a possibility distribution is described, insuring that the resulting possibility distribution will dominate the true probability distribution in at least $100(1 - \alpha)$ of the cases.
We will also use a procedure of computing possibilities for data sample, in the case where we have at hand the probability distributions of generating the data sample inside a cluster. This second kind of possibility distribution helps to label a cluster as normal or abnormal.

To build a possibility measure related to a cluster, we consider the parameter vector $p = (p_1, p_2, \ldots, p_K)$ of probabilities characterizing the unknown probability distribution of a random variable $X$ on $\Omega = \{\omega_1, \ldots, \omega_K\}$. Let $n_k$ denotes the number of observations of cluster $k$ in a sample of size $N$. Then, the random vector $n = (n_1, \ldots, n_K)$ can be considered as a *multinomial* distribution with parameter $p$. A confidence region for $p$ at level $1-\alpha$ can be computed using *simultaneous confidence intervals* as described in [4]. Such a confidence region can be considered as a set of probability distributions.

A consistency principle between probability and possibility was first stated by Zadeh [5] in an unformal way: "*what is probable should be possible*". This requirement is translated via the inequality:

$$P(A) \leq \Pi(A) \qquad \forall A \subseteq \Omega \tag{14}$$

where $P$ and $\Pi$ are, respectively, a probability and a possibility measure on a domain $\Omega = \{\omega_1, \ldots, \omega_K\}$. In this case, $\Pi$ is said to dominate $P$. Transforming a probability measure into a possibilistic one then amounts to choosing a possibility measure in the set $\Im(P)$ of possibility measures dominating $P$. This should be done by adding a strong order preservation constraint which ensures the preservation of the shape of the distribution:

$$p_i < p_j \Leftrightarrow \pi_i < \pi_j \qquad \forall i, j \in \{1, \ldots, K\}, \tag{15}$$

where $p_i = P(\{\omega_i\})$ and $\pi_i = \Pi(\{\omega_i\})$, $\forall i \in \{1, \ldots, K\}$. It is possible to search for the most specific possibility distribution verifying (14) and (15) (a possibility distribution $\pi$ is more specific than $\pi^{'}$ if $\pi \leq \pi^{'}, \forall i$). The solution of this problem exists, is unique and can be described as follows. One can define a strict partial order $\mathsf{P}$ on $\Omega$ represented by a set of compatible linear extensions $\Lambda(\mathsf{P}) = \{l_u, u = 1, L\}$. To each possible linear order $l_u$ , one can associate a permutation $\sigma_u$ of the set $\{1, \ldots, K\}$ such that:

$$\sigma_u(i) < \sigma_u(j) \Leftrightarrow (\omega_{\sigma_u(i)}, \omega_{\sigma_u(j)}) \in l_u, \tag{16}$$

The most specific possibility distribution, compatible with $p = (p_1, p_2, \ldots, p_K)$, can then be obtained by taking the maximum over all possible permutations:

$$\pi_i = \max_{u=1,L} \sum_{\{j|\sigma_u^{-1}(j)\leq\sigma_u^{-1}(i)\}} p_j \tag{17}$$

The permutation $\sigma$ is a bijection and the reverse transformation $\sigma^{-1}$ gives the rank of each $p_i$ in the list of the probabilities sorted in the ascending order. The number of permutations $L$ depends on the duplicated $p_i$ in $p$. It is equal to 1 if there is **no duplicate** $p_i$**,** $\forall i$ and for this case $\mathsf{P}$ is a *strict linear order* on $\Omega$.

In the case of searching possibilities for the data cluster themselves, we do not know the probabilities $p$, and then we aim to build *confidence intervals* for each of the cluster $c_i$. In interval estimation, a scalar population parameter is typically estimated as a range of possible values, namely a confidence interval, with a given confidence level $1 - \alpha$.

To construct confidence intervals for multinomial proportions, it is possible to find simultaneous confidence intervals with a joint confidence level $1 - \alpha$. The method attempts to find a confidence region $\mathcal{C}_n$ in the parameter space $p = (p_1, \ldots, p_K) \in [0;1]^K | \sum_{i=1}^{K} p_i = 1$ as the Cartesian product of $K$ intervals$[p_1^-, p_1^+] \ldots [p_K^-, p_K^+]$ such that we can estimate the coverage probability with:

$$\mathbb{P}(p \in \mathcal{C}_n) \geq 1 - \alpha \tag{18}$$

We can use the Goodman formulation in a series of derivations to solve the problem of constructing the simultaneous confidence intervals [6]. Let

$$A = \chi^2(1 - \alpha/K, 1) + N \tag{19}$$

where $\chi^2(1 - \alpha/K, 1)$ denotes the quantile of order $1 - \alpha/K$ of the chi-square distribution with one degree of freedom, and $N = \sum_{i=1}^{K} n_i$ denotes the size of the sample. We have also the following quantities:

$$B_i = \chi^2(1 - \alpha/K, 1) + 2n_i, \tag{20}$$

$$C_i = \frac{n_i^2}{N}, \tag{21}$$

$$\Delta_i = B_i^2 - 4AC_i. \tag{22}$$

Finally, the bounds of the confidence intervals are defined as follows:

$$[p_i^-, p_i^+] = \left[\frac{B_i - \Delta_i^{\frac{1}{2}}}{2A}, \frac{B_i + \Delta_i^{\frac{1}{2}}}{2A}\right] \tag{23}$$

It is now possible, based on these above interval-valued probabilities, to compute the most possibility distributions of the data inside a cluster, dominating any particular probability measure. Let $\mathsf{P}$ denotes the partial order induced by the intervals $[p_i] = [p_i^-, p_i^+]$:

$$(\omega_i, \omega_j) \in \mathsf{P} \Leftrightarrow p_i^+ < p_j^- \tag{24}$$

As explained above, this partial order may be represented by the set of its compatible linear extensions $\Lambda(\mathsf{P}) = \{l_u, u = 1, L\}$, or equivalently, by the set of the corresponding permutations$\{\sigma_u, u = 1, L\}$. Then for each possible permutation $\sigma_u$ associated to each linear order in $\Lambda(\mathsf{P})$, and each cluster $\omega_i$, we can solve the following linear program:

$$\pi_i^{\sigma_u} = \max_{p_1,\ldots,p_K} \sum_{\{j|\sigma_u^{-1}(j)\leq\sigma_u^{-1}(i)\}} p_j \tag{25}$$

under the constraints:

$$\begin{cases} \sum_{i=1}^{K} p_i = 1 \\ p_k^- \leq p_k \leq p_k^+ \qquad \forall k \in \{1, \ldots, K\} \\ p_{\sigma_u(1)} \leq p_{\sigma_u(2)} \leq \cdots \leq p_{\sigma_u(K)} \end{cases} \tag{26}$$

Then, we can take the distribution of the cluster $c_i$ dominating all the distributions $\pi^{\sigma_u}$:

$$\pi_i = \max_{u=1,L} \pi_i^{\sigma_u} \qquad \forall i \in \{1, \ldots, K\} \tag{27}$$

Finally we propose to build a measure of possibility distribution $\pi_{normal}$ as a threshold, and then a cluster will be considered as normal if its possibility distribution satisfies :

$$\pi_i \geq \pi_{normal}, \tag{28}$$

Otherwise it is ranged in sub space potentially suspicious. Our attention will be placed in this sub space for anomaly detection. To find the possibility distribution $\pi_{normal}$, we use the *a posteriori* probability of each gaussian component (cluster) $k$ [26]:

$$\Pr(k|x_t, \theta) = \frac{\pi_k \phi(x_t|\mu_k, \sum_k)}{\sum_{n=1}^{K} \pi_n \phi(x_t|\mu_n, \sum_n)} \tag{29}$$

which gives us, for each data point $x_t$ the probability distribution $p = (p_1, p_2, \ldots, p_K)$ (for each data point the constraints $\sum_{i=1}^{K} p_i = 1$ is always obtained from (29)).

Thereafter, we can use (17) to calculate the corresponding possibility distribution of each data point $x_t$ of the sample $x$. We obtain a matrix $\pi_K^N$ of dimension $K \times N$ (remember $K$ is the number of clusters and $N$ is the length of the data sample $x$). We take the **max** for each column (each column containing the possibility distribution for data point $x_t$). Then we obtain a second matrix $\pi_1^N$ and finally we use (30) to derive the threshold $\pi_{normal}$ :

$$\pi_{normal} = max(\pi_1^N) \tag{30}$$

## IV. Model Evaluation

### A. Experimental data: Abilene and SWITCH networks

In this work, we used a collection of data coming from the Abilene network and SWITCH one. The Abilene backbone has 11 Points of Presence(PoP) and spans the continental US. The data from this network was collected from every PoP at the granularity of IP level flows. The Abilene backbone is composed of Juniper routers whose traffic sampling feature was enabled. Of all the packets entering a router, $1\%$ are sampled at random. Sampled packets are aggregated at the 5-tuple IP-flow level and aggregated into intervals of 10 minute bins. The raw IP flow level data is converted into a PoP-to-PoP level matrix using the procedure described in [7]. Since the Abilene backbone has 11 PoPs, this yields a traffic matrix with 121 OD flows. Each traffic matrix element corresponds to a single OD flow, however, for each OD flow we have a seven week long time series depicting the evolution (in 10 minute bin increments) of that flow over the measurement period. All the OD flows have traversed 41 links. Synthetic anomalies are injected into the OD flows by the methods described in [7], and this resulted in 97 detected anomalies in the OD flows. The anomalies injected in the Abilene data are small and high *synthetic volume anomalies*. We used exactly the same Abilene data as in [14]. So for a full understanding on how the **ground-truth** is obtained (based on EWMA and Fourier algorithms) , we refer the reader to [14].

The second collection of data we used for our experiments is a set of three weeks of Netflow data coming from one of the peering links of a medium-sized ISP (SWITCH, AS559). Anomalies in the data were identified using available manual labelling methods: visual inspection of time series and top-n queries on the flow data. This resulted in 28 detected anomalous events in UDP and 73 detected in TCP traffic. We refer the reader to [8][22] for a full view of this data set.

### B. Validation

To validate our approach, we first run the MOESP algorithm in order to find the LTISS parameters and thereafter we perform a Kalman filter to perform entropy reduction and to retrieve the innovation process [23]. Thereafter the *unidimensional decision variable* is built as explained in Section III-A. As a second step, we calibrate a gaussian mixture model, using the HDDC approach, for the purpose of clustering the multivariate innovation process. The use of gaussian mixture models seems to be relevant if we assume that the innovation process is a mixture of normal distributions, instead of a simple uncorrelated gaussian white noise. It is important to note that the HDDC clustering operation is done on the multidimensional innovation matrix and not on the unidimensional decision vector, but we aim to clustering this univariate process. The HDDC method gives as output a single vector consisting of the unique sequence (class label) of symbols (alphabet) making possible to know the length of each cluster and the data belonging to it. We have used this class label to do the clustering of the decision variable. To validate our clustering model, we run the HDDC clustering operation for a set of $r$ components ($r \in \{2, \dots, 9\}$) and we select the model with the lowest value of the BIC.

The first result is about the calibration of the GMM model. In this unsupervised clustering technique, we adopt the following method to find the best number of clusters. We consider 8 partitions with different number of clusters $K \in \{2, 3, \dots, 9\}$. Since each GMM model is characterized by the mean, prior and variance vectors, the best partition is simply the one with the lowest variance vector. In our experiments, we have found $K = 3$ clusters, both for the Abilene and the UDP traffic, and $K = 4$ classes for the TCP traffic. The rest of the computations accounts for the calculations of the possibility distributions for the clusters and the data sample. In Table I we show the degree of possibility and the sample size of each cluster. To decide if a cluster in normal or not, we consider the results in Table II showing the posterior probability distributions of the data sample (given by equation (29)) easily obtained as output of the HDDC clustering, and the corresponding possibility distributions computed via equation (17), respectively for the Abilene, TCP and UDP traffic. Table II shows clearly that possibility distributions measures dominate probability distributions. So with the framework of possibility theory, we could reinforce methodology based on bayesian inference. At this point, we can easily derive the critical possibility distribution $\pi_{normal}$, calculated via (30), which is used to determine the normal clusters. The table is truncated because $N \in \{480; 1008\}$, and they show that there's for each data point (at time $t$) a cluster for which the possibility distribution is equal to 1. Then we obtain $\pi_{normal} = 1$ if we apply equation (30). Finally, a cluster $i$ will be considered as normal if its possibility distribution $\pi_i^S$ satisfies $\pi_i \geq 1$ as defined in (28). Now, if one applies equation (27), he/she obtains for the Abilene case, the vector $\{1.0000; 0.0595; 0.0465\}$ corresponding respectively to the possibility distribution of generating the clusters #1, #2 and #3 in that order. Finally, it becomes clear that, only the cluster #1 defines the normal behavior and the remaining ones are in the abnormal domain. The same reasoning performed on the SWITCH data, gives that the clusters #1 and #3 define the normal space for the TCP traffic, while the clusters #3 defines the normal behavior for the UDP traffic. It is interesting to observe that the length of clusters, belonging to the normal subspace, is always the highest, and contains most of the data. This seams to be the normal situation in anomaly detection, since anomalies might be rare and might appear in some clusters with few data. Finally to perform the detection issue, one has just to extract, from the decision variable all the points corresponding to the data belonging to the abnormal subspace (i.e., clusters labelled as suspicious), and apply thresholding (a limit strictly superior to zero) to identify and detect the anomalous events.

We have chosen in this work, as a criterion of performance, to analyze the trade-off between the false positive rate (FPR) and the detection rate (DR). The results are shown in the ROC curves in Figure 1. Typically, the natural way to analyze a ROC curve is to calculate the area under the curve. If the area is high, it means that the DR is high (approaching 100%) and the FPR low (approaching 0%). However, there are other possibles interpretations of the ROC curve. For example, one can put the x-axis in logarithmic form in order to find different points for comparison of different curves. Then, from the results depicted in Figure 1, we can see obviously that the technique based possibility distribution performs best. On can extract reference

TABLE I
POSSIBILITY DISTRIBUTIONS $\pi_i^S$ AND LENGTH OF EACH CLUSTER.

**Abilene**

| cluster $i$ | 1 | 2 | 3 | $\alpha$ |
|---|---|---|---|---|
| $p_i^-$ | 0.0424 | 0.9167 | 0.0018 | 0.05 |
| $p_i^+$ | 0.0777 | 0.9534 | 0.0137 | |
| $\pi_i^S$ | **1.0000** | *0.0595* | *0.0465* | |
| Length cluster $i$ | **936** | *50* | *22* | |

**Switch UDP**

| cluster $i$ | 1 | 2 | 3 | $\alpha$ |
|---|---|---|---|---|
| $p_i^-$ | 0.2960 | 0.0932 | 0.4746 | 0.05 |
| $p_i^+$ | 0.3993 | 0.1656 | 0.5830 | |
| $\pi_i^S$ | 0.5254 | *0.1656* | **1.0000** | |
| Length cluster $i$ | 166 | *60* | **254** | |

**Switch TCP**

| cluster $i$ | 1 | 2 | 3 | 4 | $\alpha$ |
|---|---|---|---|---|---|
| $p_i^-$ | 0.3058 | 0.0196 | 0.3337 | 0.1754 | 0.05 |
| $p_i^+$ | 0.4145 | 0.0631 | 0.4441 | 0.2693 | |
| $\pi_i^S$ | **1.0000** | *0.0631* | **1.0000** | *0.3325* | |
| Length cluster $i$ | **172** | *17* | **186** | *105* | |

TABLE II
PROBABILITIES DISTRIBUTIONS OF THE DATA SAMPLE (DECISION VARIABLE) AND CORRESPONDING POSSIBILITY DISTRIBUTIONS, ($\alpha = 0.05$).

**Abilene**

| time $t$ | 1 | 2 | 3 | ... | 1007 | 1008 |
|---|---|---|---|---|---|---|
| posterior probability distributions | | | | | | |
| *cluster*1 | 0.0353 | 0.9995 | 0.1582 | ... | 0.3908 | 0.1805 |
| *cluster*2 | 0.9647 | 0.0004 | 0.0001 | ... | 0.0001 | 0.0001 |
| *cluster*3 | 0.0000 | 0.0001 | 0.8417 | ... | 0.6091 | 0.8194 |
| possibility distributions | | | | | | |
| *cluster*1 | 0.0353 | **1.0000** | 0.1583 | ... | 0.3909 | 0.1806 |
| *cluster*2 | **1.0000** | 0.0005 | 0.0001 | ... | 0.0001 | 0.0001 |
| *cluster*3 | 0.0000 | 0.0001 | **1.0000** | ... | **1.0000** | **1.0000** |

**Switch TCP**

| time $t$ | 1 | 2 | 3 | ... | 479 | 480 |
|---|---|---|---|---|---|---|
| posterior probability distributions | | | | | | |
| *cluster*1 | 0.0000 | 0.0000 | 0.0008 | ... | 0.0000 | 0.9566 |
| *cluster*2 | 1.0000 | 1.0000 | 0.0000 | ... | 0.0000 | 0.0001 |
| *cluster*3 | 0.0000 | 0.0000 | 0.9991 | ... | 0.0000 | 0.0039 |
| *cluster*4 | 0.0000 | 0.0000 | 0.0000 | ... | 1.0000 | 0.0394 |
| possibility distributions | | | | | | |
| *cluster*1 | 0.0001 | 0.0000 | 0.0011 | ... | 0.0001 | **1.0000** |
| *cluster*2 | **1.0000** | **1.0000** | 0.0000 | ... | 0.0002 | 0.0001 |
| *cluster*3 | 0.0000 | 0.0003 | **1.0000** | ... | 0.0001 | 0.0041 |
| *cluster*4 | 0.0002 | 0.0001 | 0.0000 | ... | **1.0000** | 0.0423 |

**Switch UDP**

| time $t$ | 1 | 2 | 3 | ... | 479 | 480 |
|---|---|---|---|---|---|---|
| posterior probability distributions | | | | | | |
| *cluster*1 | 0.0000 | 0.0000 | 0.6989 | ... | 0.0000 | 0.1707 |
| *cluster*2 | 1.0000 | 1.0000 | 0.0000 | ... | 0.0000 | 0.8041 |
| *cluster*3 | 0.0000 | 0.0000 | 0.3011 | ... | 1.0000 | 0.0251 |
| possibility distributions | | | | | | |
| *cluster*1 | 0.0001 | 0.0001 | **1.0000** | ... | 0.0002 | 0.2918 |
| *cluster*2 | **1.0000** | **1.0000** | 0.0003 | ... | 0.0001 | **1.0000** |
| *cluster*3 | 0.0002 | 0.0000 | 0.5102 | ... | **1.0000** | 0.0312 |

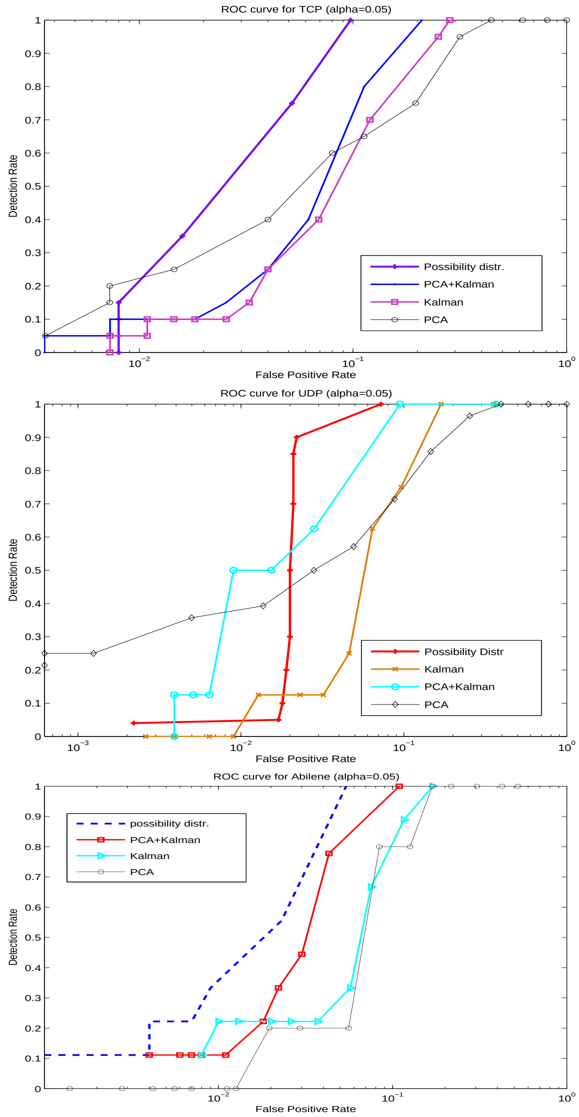


Fig. 1. ROC curve for the DR vs FPR. Top left graph for TCP, top right graph for UDP and top down for Abilene data when $\alpha = 0.05$.

points for which the FPR decrease significantly for our new scheme than for the other three techniques we had already derived in our previous works [23][27][28]. The method shows that we can achieve a DR of 100% with a FPR equal to 5%, where the best method of the three others, namely the PCA-Kalman method exhibits a FPR equal to 10%, for the Abilene data. For the SWITCH data, the new approach can achieve a probability of detection of 90% with a FPR of 2% against 7%, for the PCA-Kalman methodology, for UDP traffic. The same situation is observed for the TCP traffic.

## V. CONCLUSION

In this work, we have shown the effectiveness and robustness of combining probability distributions, and possibility distributions for the purpose of anomaly detection. The robustness of the approach is achieved, in part, by the use of subspace identification algorithms (via the aid of PCA) and Kalman filtering technique, in order to build a unidimensional decision variable from multidimensional data set. Moreover, the great innovation in this paper is the use of possibility distributions to find the normal behavioral model, (by means of simple transformations from probability distributions) allowing us to extract the anomaly space. Another benefit of the solution can be found in the simplicity of the all procedure and the low complexity making easy to implement the algorithm. On the other hand, we have performed a robust and efficient high dimensional data clustering to build normal clusters with most of the data, and abnormal ones containing a few number of data where all anomalies lie. The experiments are done on different real traffic, and the ROC curve has shown high performance, compared to other techniques. It seems the main drawback of this work comes from the fact that the final decision process is based on applying manual thresholding. This problem will thus limit the applicability of the solution to dynamic and evolving systems. It will be of interest to search for more convenient technique, to automatically and dynamically adjust this threshold. We will try to address this issue soon.

## REFERENCES

[1] Dubois D., Prade, H. and Sandri, S.: On possibility/probability transformations. In Proceedings of the Fourth Int. Fuzzy Systems Association World Congress (IFSA91), Brussels, Belgium, 1991, pages 50-53.

[2] Dubois, D. Foulloy, L., Mauris, G. and Prade, H. Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities. Reliable Computing, 2004, pp. 273-297.

[3] Dubois, D., Prade, H. and Sandri, S. On possibility/probability transformations. In Proceedings of the Fourth Int. Fuzzy Systems Association World Congress (IFSA91), Brussels, Belgium, 1991, pages 50-53.

[4] Masson, M., H. and Denoeux, T.: Inferring a possibility distribution from empirical data. Fuzzy Sets and Systems 157(3), 2006, pp. 319-340.

[5] Zadeh, L.,A. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1978, pp. 3-28.

[6] Goodman, L. A.: On simultaneous confidence intervals for multinomial proportions. Technometrics, 7(2): 1965, pp. 247-254.

[7] Lakhina, A., Crovella, M. and Diot, C.: Characterization of network-wide traffic anomalies. In Proceedings of the ACM/SIGCOMM Internet Measurement Conference, 2004. pp. 201-206.

[8] Brauckhoff, D., Salamatian, K. and May, M.: Applying PCA for Traffic Anomaly Detection: Problems and Solutions. Proceedings IEEE INFOCOM, 2009 pp. 2866-2870.

[9] Ringberg, H.,Soule, A., Rexford, J., and Diot, C.:Sensitivity of PCA for Traffic Anomaly Detection. In ACM SIGMETRICS 2007.

[10] Soule, A., Salamatian, K.,Taft, N.: Traffic Matrix Tracking using Kalman Filters. ACM LSNI Workshop 2005.

[11] Soule, A., Salamatian, K. and Taft, N.: Combining Filtering and Statistical Methods for Anomaly Detection. USENIX , Association, Internet Measurement Conference, 2005, pp. 331344.

[12] Shumway, R. H. and Stoffer, D. S.: Dynamic Linear Models With Switching. Journal of the American Statistical Association, 1992, pp. 763-769.

[13] Eriksson, B., Barford, P., Bowden, R., Duffield, N., Sommers, J., Roughan, M. : BasisDetect: a model-based network event detection framework. In ACM IMC 2010.

[14] Lakhina, A., Crovella, M.,Diot, C.: Diagnosing Network-Wide Traffic Anomalies. In ACM SIGCOMM 2004.

[15] Brauckhoff, D., Salamatian, K. and May, M: Applying PCA for Traffic Anomaly Detection: Problems and Solutions. Technical Report INFOCOM 2009.

[16] Katayama, T.: subspace Methods for System Identification, Springer 2005.

[17] Verhaegen, M.: Identification of the Deterministic part of MIMO State Space Models given in Innovations Form from Input-Output Data. Journal Automatica, 1994, vol. 30 No 1.pp 61-74.

[18] Bottura, C.P, Tamariz, A.D.R, Barreto, G. and Cáceres, A.F.T: Parallel and Distributed MOESP Computational system's Modelling. Proceedings of the 10th Mediterranean Conference, 2002.

[19] Shumway, R. H. and Stoffer, D. S.: An Approach to Time Series Smoothing And Forecasting Using the EM Algorithm. Journal of Time Series Analysis, 1982, vol.3, No 4.

[20] Cattell, R.: The scree test for the number of factors. Multivariate Behavioral Research, 1966, pp. 245-276.

[21] Kailath, T., Sayed, A. H. and Hassibi B.: Linear Estimation. Prentice Hall, 2000.

[22] Brauckhoff, D., Dimitropoulos, X., Wagner, A. and Salamatian, K. : Anomaly Extraction in Backbone Networks using Association Rules. IMC09, November 46, Chicago, Illinois, USA, 2009.

[23] Ndong, J., Salamatian, K., :A Robust Anomaly Detection Technique Using Combined Statistical Methods. CNSR 2011, *IEEE Xplore* 978-1-4577-0040-8, 2011, pp: 101-108.

[24] Schwarz, G.: Estimating the dimension of a model. Annals of Statistics, 1978, PP 461-464.

[25] Douglas A. R.: Gaussian Mixture Models. Encyclopedia of Biometrics, 2009, pp. 659-663.

[26] Bouveyron, C.,Girard, S., and Schmid, C. High-Dimensional Data Clustering, Computational Statistics and Data Analysis, vol. 52 (1), 2007, pp. 502-519.

[27] Ndong, J., Salamatian, K.,: Signal Processing-based Anomaly Detection Techniques: A Comparative Analysis. INTERNET 2011, The Third International Conference on Evolving Internet. ISBN: 978-1-61208-141-0.

[28] Ndong, J.,: Anomaly Detection: A Technique Using Kalman Filtering and Principal Component Analysis. ATAI NTC 2012 GSTF 2012.

# Quality of Service Assurance in Multi-domain Content-Aware Networks for Multimedia Applications

Marius Vochin, Eugen Borcoci, Serban Georgica Obreja, Cristian Cernat, Radu Badea, Vlad Poenaru
University POLITEHNICA of Bucharest
Bucharest, Romania
emails: marius.vochin@elcom.pub.ro, eugen.borcoci@elcom.pub.ro
serban.obreja@elcom.pub.ro, cristian.cernat@elcom.pub.ro, radu.badea@elcom.pub.ro, vlad.poenaru@elcom.pub.ro

***Abstract* —Content Aware Networking is a "light" variant of Information Centric Networking (ICN) architectural solution, responding to the significant Internet orientation to content and multimedia. The contribution of this paper is focused on the implementation and new experimental results validating previously designed solutions for Virtual Content Aware Networks (VCAN), QoS enabled, built over multi-domain, multi-provider IP networks, in the framework of an ecosystem offering multimedia services. A multi-domain pilot is shortly described, and QoS related experiments are shown, demonstrating the benefit of provisioned VCANs dedicated to transport media flows with QoS guarantees.**

***Keywords — Content-Aware Networking, Network Aware Applications, Multi-domain, Quality of Services, Multimedia distribution, Future Internet.***

## I. Introduction

A significant Internet trend is its stronger information/content-centric orientation [1]-[5]. Related to this, a high increase in media traffic amount is seen, driven by media oriented applications. On the other side, several current Internet limitations have been emphasized, related to the needs of today communication. Consequently, changes in services and networking have been proposed, including modifications of the architectural basics. The *Information/Content-Centric Networking (ICN/CCN)*, [3][4], approaches propose revision of some main concepts of the architectural TCP/IP stack. In parallel, evolutionary solutions emerged, as Content-Awareness at Network layer (CAN) and Network-Awareness at Applications layers (NAA), seen as a light ICN solution. This approach can create a powerful cross-layer optimisation loop between the transport, applications and services.

The European FP7 project, "Media Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments", ALICANTE, [10]–[13], adopted the NAA/CAN approach. It defined a multi-domain architecture, and then specified, designed and implemented a Media Ecosystem, on top of multi-domain IP networks, to offer media services for business actors playing roles of consumers and/or providers. The solution adopted as main principles the content-type recognition at network level and light virtualization (separated Data Plane virtual networks, but a single management and control plane). This solution is believed to offer seamless deployment perspectives and tries to avoid the scalability problems, while are still open research issues, of the full ICN/CON approaches.

In this paper, the "environment" denotes a generic grouping of functions working for a common goal, which vertically span one or more several architectural (sub)layers. Several cooperating environments are defined, including business entities/actors: *User Environment (UE),* containing the End-Users; *Service Environment (SE),* containing High Level Service Providers (SP) and Content Providers (CP); *Network Environment (NE),* where a new CAN Provider exists (CANP - managing and offering Virtual Content Aware Networks- VCANs); traditional Network Providers (NP/ISP) - managing the network elements at IP level.

A VCAN is constructed at request of an SP, to a CAN Provider, based on Service Level Agreement (SLA) spanning one or several independent network domains, and featuring different levels of QoS guarantees. The CANP offers to the upper layers enhanced connectivity services, VCAN-based QoS enabled, in unicast and multicast mode, over multi-domain, multi-provider IP networks. The VCAN resources are managed quasi-statically by provisioning and also dynamically, by using adaptation procedures for media flows. In the Data Plane, content/service description information (metadata) can also be inserted in the media flow packets by the Content Servers and treated appropriately by the intelligent routers of the VCAN.

This paper continues a previous work dedicated to system architecture, design, and algorithms for resource management. Section II makes a short overview of related work. Section III summarizes the overall system architecture. Section IV and V contain the main novel contributions of this paper, presenting the experimental pilot structure and few samples of validation results, extracted of a large set, [12] performed during overall system validation. Section VI contains some conclusions and future work outline.

## II. Related Work

The current challenges and research on content/information networking for the Future Internet are well presented in [1][2]. Information/Content Centric Networking (ICN/CCN), or Content Oriented Networking (CON) [3]–[5] change the traditional TCP/IP stack concepts (with agnostic network layer) by putting more intelligence in the network nodes, which primarily process the data, based on *content objects* recognition and not based on *location address*. However, full ICN/CCN poses significant problems of scalability and seamless deployment possibility [6].

"Light" approaches are Content-Awareness at Network layer (CAN) and Network-Awareness at Applications layers (NAA) [10]. Here, only content-type of data are recognized by network nodes and treated accordingly. Thus the full ICN scalability and complexity problems are avoided and a. seamless field deployment is possible. The approach brings new benefits for both, Service and Application Layer and Network layer, thus creating a powerful *cross-layer optimization loop*.

Network virtualization is an important set of tools to overcome the ossification of the current Internet [7]-[9]. In our system [10]–[13], a light virtualization is adopted, by creating VCANs as parallel Internet planes [14], each one being content aware and QoS capable. The virtual logical slices (VCANs) exist and are separated in the Data Plane only, while the Management and Control (M&C) Planes are unique at system level. To solve the QoS guarantees, each VCAN has associated a QoS class or meta-QoS class as in [15][16], but in a more powerful way, given the content awareness of the VCANs.

## III. ALICANTE System Architecture Summary

### A. General Architecture

The general ALICANTE architecture is already defined in [10][11][13]. The traditional business actors are SP, CP, NP - Providers and End-Users (EU). A novel actor is the CAN Provider (CANP) which offers virtual layer connectivity services. The EUs are connected to VCANs via home and/or access networks. Note that VCANs do not control access networks resources, given their large variety of technologies. Home-Boxes (HB) – may exist, partially managed by the SP, the NP, and the end-user, located at end-user's premises and gathering content/context-aware and network-aware information. The HB can also act as a CP/SP for other HBs, on behalf of the EUs. Correspondingly, two novel virtual layers exist: the CAN layer and the HB layer.

The novel CAN routers are called *Media-Aware Network Elements (MANE)* to emphasize their additional capabilities: content and context - awareness, controlled QoS/QoE, security and monitoring features, etc.

The CAN layer M&C is partially distributed. It supports CAN customization to respond to the high level services needs, including 1:1, 1:n, and n:m communications and also performs efficient network resource exploitation. The interface between CAN and the upper layer assures *cross-layer optimizations* interactions. A hierarchical monitoring subsystem (at user, service, CAN and respectively network layer) supervises several points of the service distribution chain and feeds the adaptation subsystems with appropriate information, at the HB and CAN Layers.

Figure 1 presents a partial view on the ALICANTE architecture (complete description is in [11]), with emphasis on the CAN layer and management interaction. The network contains several Core Network Domains (CND), belonging to NPs (they can be also seen as Autonomous Systems - AS). The Access Networks ANs are out of scope of VCANs. One *CAN Manager* (CANMgr) exists for each CAN domain to assure the consistency of VCAN planning, provisioning, advertisement, offering, negotiation installation and exploitation. Each domain has an *Intra-domain Network Resource Manager* (IntraNRM), as the ultimate authority configuring the network nodes. The CAN layer cooperates with HB and SE and offers CAN services to them.

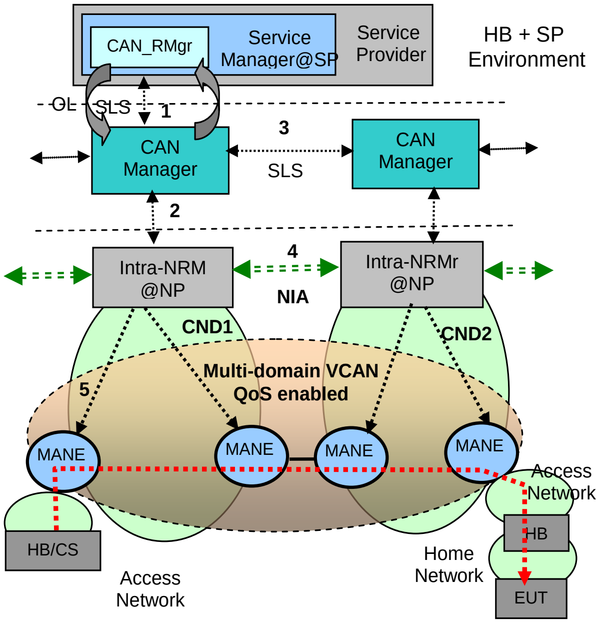


Figure 1: High level ALICANTE architecture: multi-domain VCANs and main management and control interactions

Notations: RM – Resource Management; HB - Home Box; CS - Content Server; EUT - End User Terminal; OL - Optimization Loop; NIA - Network Interconnection Agreements; SP, NP – Service, Network Providers

### B. VCAN Management

The VCAN Management framework has been defined in [13]. *A short summary is recalled here only for sake of clarity.* At the Service Manager SM@SP, the CAN Network Resources Manager (CAN_RMgr) performs, on behalf of SP, all actions needed for VCAN support: *VCAN* planning, provisioning (i.e. negotiation with CANP) and then VCAN operation supervision. The *CANMgr@CANP* performs, at the CAN layer, VCAN provisioning and operation. The two entities interact based on the SLA/SLS contract initiated by the SP. The interface implementation for management is based on Simple Object Access Protocol (SOAP)/Web Services.

The M&C contracts/interactions of SLA/SLS types (Figure 1) are: *SP-CANP(1)*: the SP requests to CANP to provision/ modify/ terminate VCANs while CANP replies with yes/no; *CANP-NP(2)*: CANP negotiates resources with NP; *CANP-CANP(3):* negotiations might be needed to extend a VCAN upon several NP domains; *Network Interconnection Agreements (NIA) (4)* between the NPs or between NPs and ANPs; the latter are not novel functionalities, but are necessary for NP cooperation. After the SP negotiates a desired VCAN with CANP, it will issue the installation commands to CANP, which in turn configures, via Intra-NRM (action 5), the MANE functional blocks (input and output).

The content awareness (CA) is realized in three ways:

(i) by concluding a SP **-** CANP SLA concerning VCAN construction. The Content Servers (CS) are instructed by the SP to insert some special *Content Aware Transport Information (CATI*) in the data packets. This simplifies the media flow classification and treatment by the MANE; (ii) SLA is concluded, but no CATI is inserted in the data packets (legacy CSs). The MANE applies packet inspection for data flow classification and assignment to VCANs. The flows treatment is still based on VCANs characteristics defined in the SLA; (iii) no SP–CANP SLA exists and no CATI. The flows treatment can still be CA, but conforming to the local policy at CANP and IntraNRM.

The DiffServ and/or MPLS technologies support splitting the sets of flows in QoS classes (QC), with a mapping between the VCANs and the QCs. Several levels of QoS granularity can be established when defining VCANs, by using one of the implemented QCs: EF, AF1, AF2 or BE. The QoS behavior of each VCAN (seen as one of the parallel Internet planes) is established by the SP-CANP.

Generally a 1-to-1 mapping between a VCAN and a network plane will exist. Customization of VCANs is possible in terms of QoS level of guarantees (weak or strong), QoS granularity, content adaptation procedures, degree of security, etc. A given VCAN can be realized by the CANP, by combining several processes, while being possible to choose different solutions concerning routing and forwarding, packet processing, and resource management.

The mapping between multiple domain VCANs and network resources are developed in [17]. Special combined novel algorithms for multi-domain QoS enabled routing, resource reservation (at aggregated level) and VCAN final mapping have been designed and implemented as to assure VCAN QoS capabilities.

## IV. EXPERIMENTAL PILOT INFRASTRUCTURE

Figure 2 presents a part of the ALICANTE pilot, i.e. an island deployed in Bucharest, UPB. It is a hybrid diagram showing both physical network infrastructure and some architectural elements.

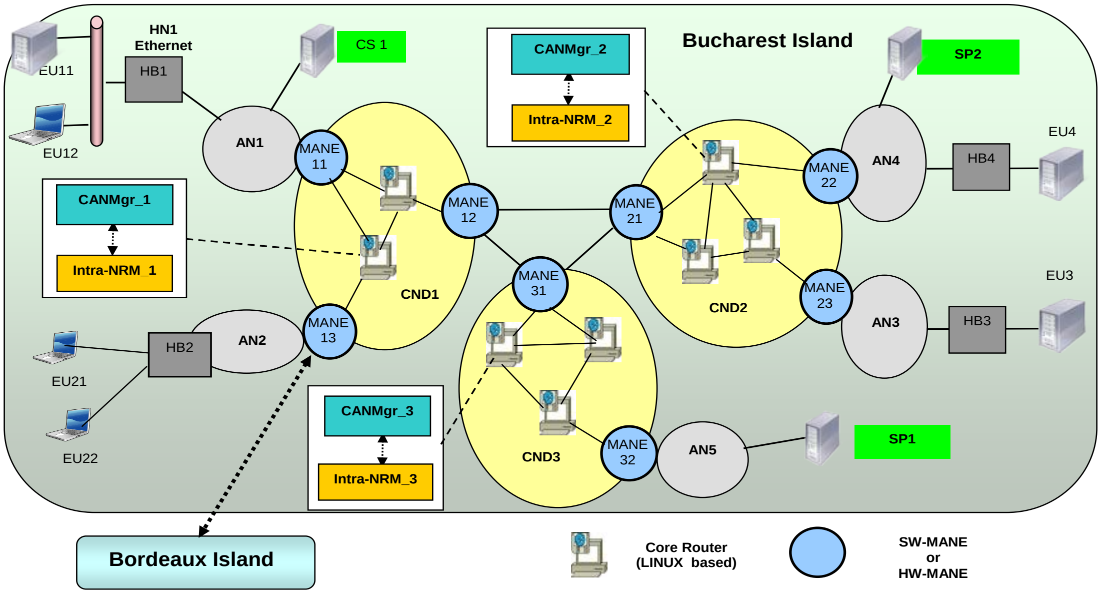


Figure 2: Bucharest Island – general network Infrastructure and architectural elements

The island is also connected via international links to other islands. Here, a summary only emphasizing some specific aspects is given. Its objectives are to validate and demonstrate the capabilities of the overall system for: CAN and network environment (CNE) transport capabilities; CAN and networking support for High-level services. Several Use Cases have been defined in the project to validate high level services capabilities: UC1 - “Distribution of User Generated Content”, UC2 - “High Popularity Video on Demand” and UC3 - “Multicast Live TV”.

The UPB island is composed by three Core Network Domains and several access networks (ANs). The infrastructure allows multi-domain and multi-provider experiments to be performed locally and also in multi-island environment (e.g., Bucharest - Bordeaux). Specifically, the pilot supported the validation of the following: CAN functionality (mono and multi-domain experiments); High-level services scenarios UC1, UC2, UC3 (local/local-remote experiments, mono and multi-domain). In the local demonstration, all business entities (EUs, HBs, SP, CP/CS entities) are placed within, and connected to the UPB island network infrastructure. In the multi-island demonstrations, a part of the service related entities resides in the UPB island and others are connected to the other pilot islands. The roles of the ALICANTE entities are detailed in [11].

The MANE edge routers are content-aware while the interior routers are core regular routers, DiffServ and MPLS capable. The core routers were implemented on Linux machines. The MANE is based on HW + SW configurations provided by the ALICANTE project. The CAN Managers and Intra-NRMs run on Linux based PC machines. The HBs, SP and CP Subsystem’s entities (SP or CP servers, Content Servers (CS), Service management entities, Service Registry, etc.) run also on Linux based PC machines. Access Network and Home networks are based on Ethernet technology.

Details on the tests performed in the project pilot, and complete sets of validation results are described in [12].

## V. VALIDATION RESULTS

This section presents a few samples of the experimental results performed to validate the system.

In particular the following example shows a VCAN created, to demonstrate its ability to protect unicast media traffic. It has two logical channels (pipes): one inter-domain pipe, starting in Domain 2 and ending in Domain 1, and a second intra-domain located in Domain 1 – see Figure 3. This test shows that traffic belonging to the VCAN pipe is protected against traffic with low priority, in this case, best effort traffic. A video stream is sent through the inter-domain VCAN's pipe, from MANE 22 to MANE 11. The bit rate for the video stream is around 2.5Mbps. A high rate data flow, streamed from node e.134 to node e.131, will be used as "noise" traffic. The data rate for the "noise" traffic is sent using *iperf* application. It is UDP traffic and it is sent with a data rate of 1Gbps. The results obtained with and without VCAN installed will be presented below.

The VCAN's pipes have a Committed Information Rate (CIR) of 4Mbps and a Peak Information Rate (PIR) of 6Mbps.

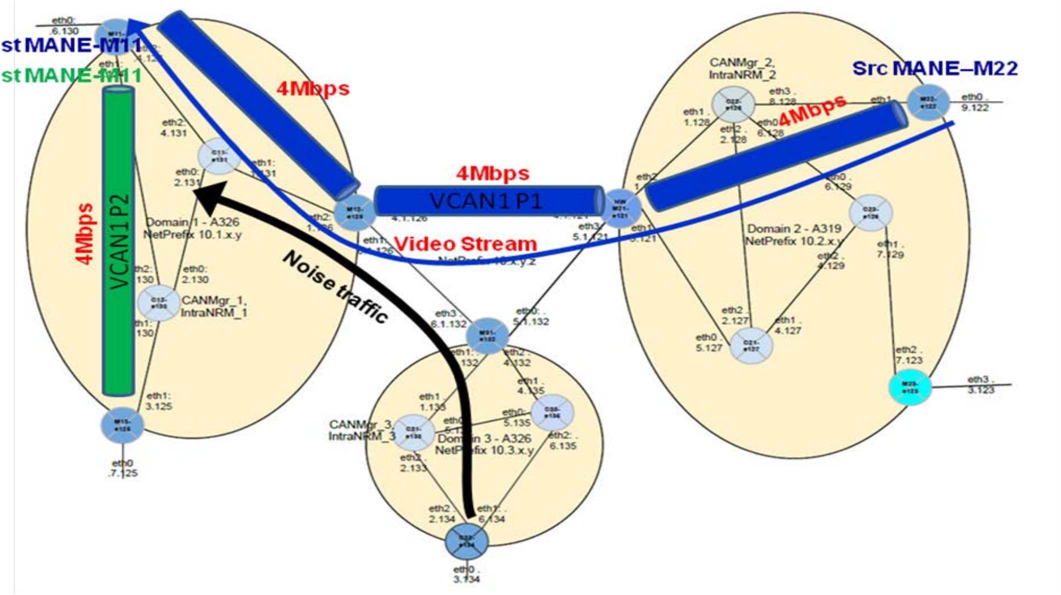
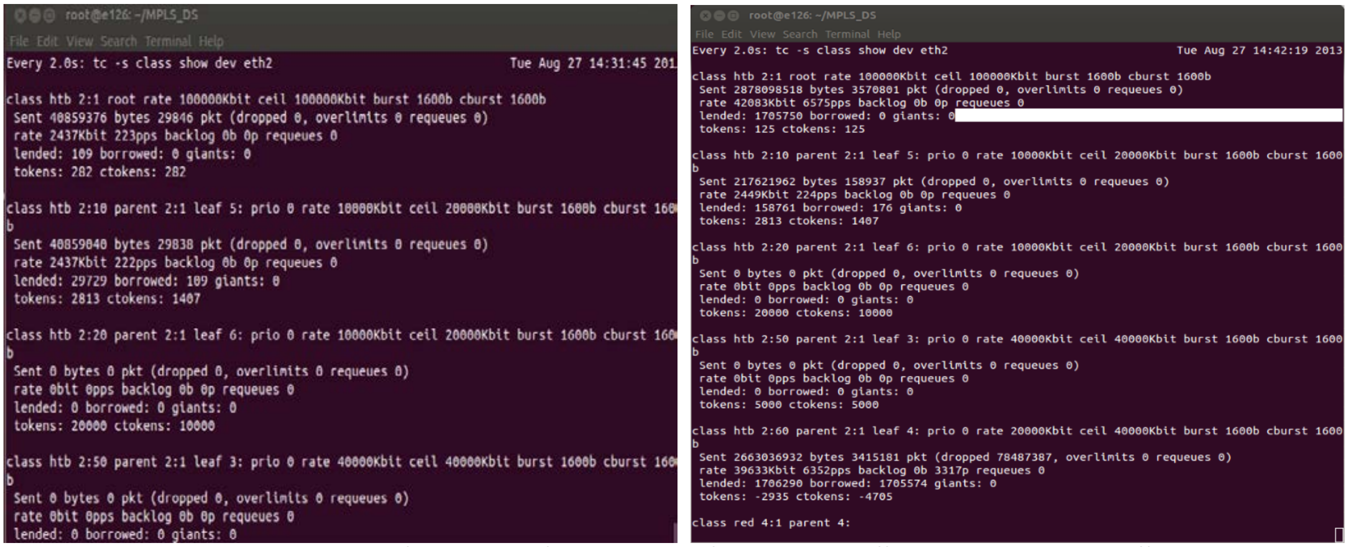


Figure 3: Sample of VCAN topology installed for unicast oriented tests.

The system has been fully implemented for its M&C and data plane components. In Figure 4, the queues of the Hierarchical Token Bucket (HTB) for Traffic Control Filters (TC) filters installed on the MANE 12 node (e.126) are shown. The HTB filters shown belong to the interface eth2, which represents the output interface for the VCAN pipe 1. Appropriate classes are created as a result of the VCAN Create request, issued by the Service Provider. In the left figure, the "noise" traffic is not present, that's why only the queue associated with the VCAN 1 traffic has traffic in it. The data rate measured for this queue is around 2.5 Mbps, which corresponds to the movie's bit rate. In the right figure, the queues belonging to the same filter are shown, but in this case, the "noise" traffic is sent through the e.126 node. It can be seen that, in the default queue for the best effort traffic, the "noise" traffic is present. Its rate is around 40Mbps, because it is reduced by the HTB filter.


root@e126: ~/MPLS_DS
File Edit View Search Terminal Help
root@e126: ~/MPLS_DS
File Edit View Search Terminal Help

```
Every 2.0s: tc -s class show dev eth2                              Tue Aug 27 14:31:45 201

class htb 2:1 root rate 100000Kbit ceil 100000Kbit burst 1600b cburst 1600b
 Sent 40859376 bytes 29846 pkt (dropped 0, overlimits 0 requeues 0)
 rate 2437Kbit 223pps backlog 0b 0p requeues 0
 lended: 109 borrowed: 0 giants: 0
 tokens: 282 ctokens: 282

class htb 2:10 parent 2:1 leaf 5: prio 0 rate 10000Kbit ceil 20000Kbit burst 1600b cburst 160
b
 Sent 40859040 bytes 29838 pkt (dropped 0, overlimits 0 requeues 0)
 rate 2437Kbit 222pps backlog 0b 0p requeues 0
 lended: 29729 borrowed: 109 giants: 0
 tokens: 2813 ctokens: 1407

class htb 2:20 parent 2:1 leaf 6: prio 0 rate 10000Kbit ceil 20000Kbit burst 1600b cburst 160
b
 Sent 0 bytes 0 pkt (dropped 0, overlimits 0 requeues 0)
 rate 0bit 0pps backlog 0b 0p requeues 0
 lended: 0 borrowed: 0 giants: 0
 tokens: 20000 ctokens: 10000

class htb 2:50 parent 2:1 leaf 3: prio 0 rate 40000Kbit ceil 40000Kbit burst 1600b cburst 160
b
 Sent 0 bytes 0 pkt (dropped 0, overlimits 0 requeues 0)
 rate 0bit 0pps backlog 0b 0p requeues 0
 lended: 0 borrowed: 0 giants: 0
```

```
Every 2.0s: tc -s class show dev eth2                              Tue Aug 27 14:42:19 2013

class htb 2:1 root rate 100000Kbit ceil 100000Kbit burst 1600b cburst 1600b
 Sent 2878098518 bytes 3570801 pkt (dropped 0, overlimits 0 requeues 0)
 rate 42083Kbit 6575pps backlog 0b 0p requeues 0
 lended: 1705750 borrowed: 0 giants: 0
 tokens: 125 ctokens: 125

class htb 2:10 parent 2:1 leaf 5: prio 0 rate 10000Kbit ceil 20000Kbit burst 1600b cburst 1600
b
 Sent 217621962 bytes 158937 pkt (dropped 0, overlimits 0 requeues 0)
 rate 2449Kbit 224pps backlog 0b 0p requeues 0
 lended: 158761 borrowed: 176 giants: 0
 tokens: 2813 ctokens: 1407

class htb 2:20 parent 2:1 leaf 6: prio 0 rate 10000Kbit ceil 20000Kbit burst 1600b cburst 1600
b
 Sent 0 bytes 0 pkt (dropped 0, overlimits 0 requeues 0)
 rate 0bit 0pps backlog 0b 0p requeues 0
 lended: 0 borrowed: 0 giants: 0
 tokens: 20000 ctokens: 10000

class htb 2:50 parent 2:1 leaf 3: prio 0 rate 40000Kbit ceil 40000Kbit burst 1600b cburst 1600
b
 Sent 0 bytes 0 pkt (dropped 0, overlimits 0 requeues 0)
 rate 0bit 0pps backlog 0b 0p requeues 0
 lended: 0 borrowed: 0 giants: 0
 tokens: 5000 ctokens: 5000

class htb 2:60 parent 2:1 leaf 4: prio 0 rate 20000Kbit ceil 40000Kbit burst 1600b cburst 1600
b
 Sent 2663036932 bytes 3415181 pkt (dropped 78487387, overlimits 0 requeues 0)
 rate 39633Kbit 6352pps backlog 0b 3317p requeues 0
 lended: 1706290 borrowed: 1705574 giants: 0
 tokens: -2935 ctokens: -4705

class red 4:1 parent 4:
```

Figure 4: The queues of the HTB TC filter on *e.126*: left - no "noise" traffic; right - with "noise" traffic

Several complex test campaigns have been performed to fully validate the functionality and measure the system performance, [12]. Given the limited space of this paper a qualitative example is given below.

In this test, two movies were continuously streamed at a constant bitrate, the left one is protected by the installed VCAN, and the right one is sent best-effort. In Figures 5 and 6, a snapshot of the videos received at the output of the VCAN pipe is shown. Both snapshots are taken with one VCAN configured and installed. Figure 5 corresponds to the case when the background noise is not present and both movies transmitted are running fine. Figure 6 is taken with background noise flowing active. The movie protected by VCAN will have good quality because its flow is prioritized and protected against low priority traffic, while the best effort movie is very affected by noise. The conclusion is that VCAN has a provisioning effect for QoS protection of dedicated flows.

## VI. CONCLUSIONS AND FUTURE WORK

The paper presented experimental results, validating the solutions for a Media Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments", developed over multiple IP network domains. The architecture and implementation has been fully validated, giving the possibility for multiple providers to cooperate by creating overlay networks over multiple domains.

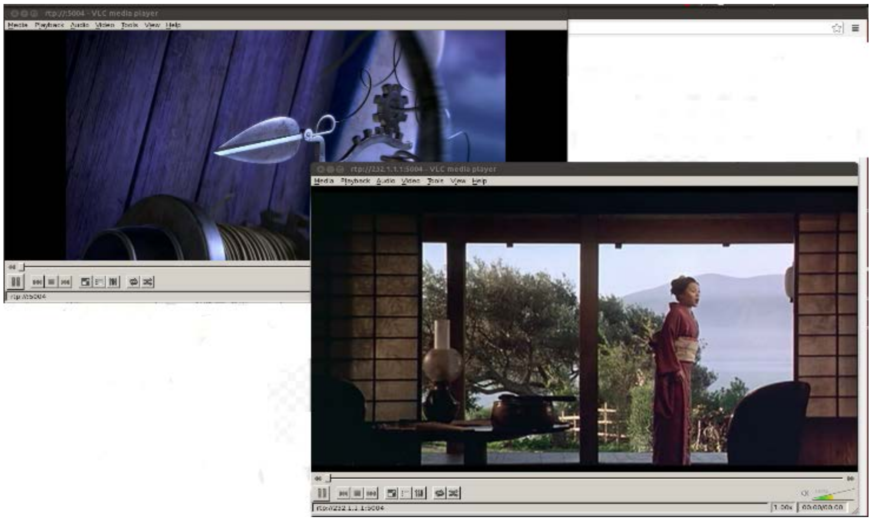

rtp://:5004 - VLC media player
Media Playback Audio Video Tools View Help
rtp://232.1.1.1:5004 - VLC media player
Media Playback Audio Video Tools View Help
rtp://232.1.1.1:5004

Figure 5: Received video stream with VCAN installed, and no background noise

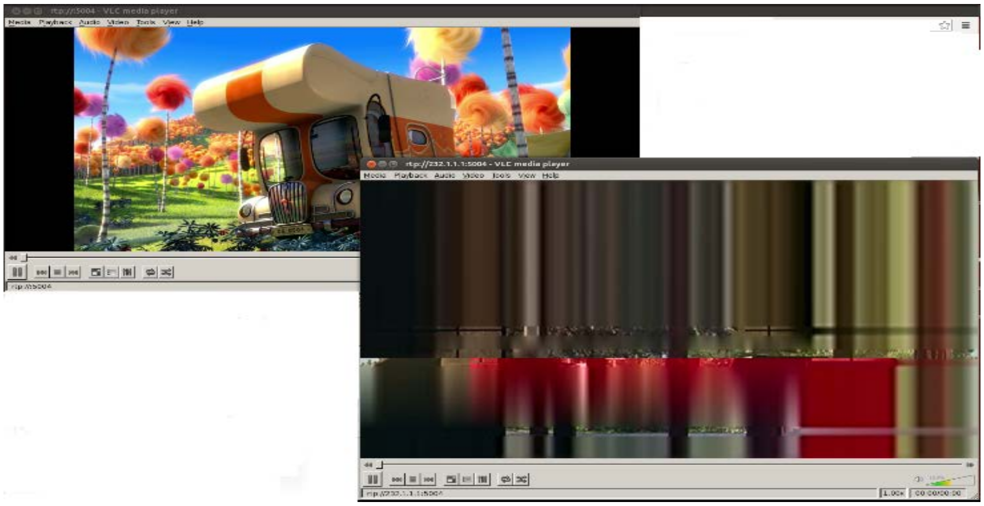

rtp://:5004 - VLC media player
Media Playback Audio Video Tools View Help
rtp://:5004
rtp://232.1.1.1:5004 - VLC media player
Media Playback Audio Video Tools View Help
rtp://232.1.1.1:5004
1.00x
00:00:00:00

Figure 6: Received video stream with VCAN installed and network loaded with background "noise" (left: protected traffic flowing through VCAN pipe; right: unprotected traffic)

We emphasize the main merit of the system, which provides the possibility of creation of parallel multi-domain VCANs customised for different classes of services, with content type awareness in edge routers, while the QoS supporting technologies can be diverse (e.g. DiffSErv, MPLS, combinations).

The VCANs, provisioned as overlays by the CAN providers, at Service Providers requests, can serve properly the media flow transportation in different conditions of the network load. The system still admits best effort traffic, thus preserving network neutrality.

Future work will be done to extend the functionality of the system towards Software Defined Networking, given that the proposed architecture fulfills the main SDN concepts – decoupling between the Control and Data Plane, partial centralization and programmability of network level forwarders.

## ACKNOWLEDGMENTS

This work was supported partially by the EC in the context of the ALICANTE project (FP7-ICT-248652) and partially by the 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, pp. 26-36, July 2011.

[2] C. Baladrón, "User-Centric Future Internet and Telecommunication Services", in: G. Tselentis, et. al. (eds.), Towards the Future Internet, IOS Press, pp. 217-226, 2009.

[3] 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.

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

[5] W. K. Chai, et. al., "CURLING: Content-Ubiquitous Resolution and Delivery Infrastructure for Next-Generation Services", IEEE Communications Magazine, March 2011, pp. 112 - 120.

[6] A. Ghodsi, S. Shenker, T. Koponen, A. Singla, B. Raghavan, and J. Wilcox, "Information-centric networking: seeing the forest for the trees", In Proc. HotNets, 2011, pp. 1:1-1:6.

[7] T. Anderson, L. Peterson, S. Shenker, and J. Turner,, "Overcoming the Internet Impasse through Virtualization", Computer, vol. 38, no. 4, pp. 34–41, Apr. 2005.

[8] 4WARD, "A clean-slate approach for Future Internet", http://www.4ward-project.eu/.

[9] N. M. Chowdhury and Raouf Boutaba**,** **"**Network Virtualization: State of the Art and Research Challenges", IEEE Communications Magazine, July 2009, pp. 20-26.

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

[11] ALICANTE, Public Deliverable D2.1, "Overall System and Components Definition and Specifications", http://www.ict-alicante.eu, Sept. 2011.

[12] ALICANTE, Public Deliverable D8.3: "Trials and Validation", http://www.ict-alicante.eu, Sept. 2013.

[13] E. Borcoci, S. Obreja, et. al., "Resource Management in Multi-Domain Content-Aware Networks for Multimedia Applications , International Journal on Advances in Networks and Services", vol 5 no 1 & 2, year 2012, pp. 43-57.

[14] M. Boucadair, et al., "A Framework for End-to-End Service Differentiation: Network Planes and Parallel Internets", IEEE Communications Magazine, Sept. 2007, pp. 134-143.

[15] P. Levis, M. Boucadair, P. Morrand, and P. Trimitzios, "The Meta-QoS-Class Concept: a Step Towards Global QoS Interdomain Services", Proc. of IEEE SoftCOM, Oct. 2004.

[16] M. P. Howarth, et al., "Provisioning for Interdomain Quality of Service: the MESCAL Approach", IEEE Communications Magazine, June 2005, pp. 129-137.

[17] R. Miruta and E. Borcoci, "Optimization of Overlay QoS Constrained Routing and Mapping Algorithm for Virtual Content Aware networks" ICNS 2013 - Lisbon, Portugal. http://www.thinkmind.org/index.php?view=article&articleid=icns_2013_4_30_10174, March, 2013.

# Edge-to-Edge Achieved Transfer Throughput Inference Using Link Utilization Counts

Demetris Antoniades and Constantine Dovrolis
School of Computer Science
College of Computing
Georgia Institute of Technology
Atlanta, Georgia
Email: [danton,constantine]@gatech.edu

***Abstract*—We propose a methodology to infer edge-to-edge achieved transfer throughput using link utilization counts. Our method treats variations in the link utilization time-series as possible transfer starting or ending events. Iteratively following these variations to the neighboring routers, we then identify the path the transfer traversed through the monitored network. Our evaluation shows that this method can identify events larger than 3 Mbit/sec and longer than 2 minutes in duration with more than 95% recall. Additionally, we show that event detection is strongly correlated with the traffic in the busiest router in the path. We discuss how a number of applications such as throughput prediction and DDoS attack source detection can use the inferred information.**

***Keywords-throughput inference; edge-to-edge; SNMP; network performance monitoring.***

## I. Introduction

The Simple Network Management Protocol (SNMP) [1], [2] is widely used to monitor aggregated link usage from network components (routers, switches, etc.). Such data, provide a valuable resource for network administrators, aiding decisions about network routing, provisioning and configuration. SNMP data is simple to collect and maintain, providing a low disk space option for a log of historical network usage.

On the other hand, Netflow data provides detailed information about end-to-end performance. Using Netflow, one can have information about the communication between two hosts, the amount of packets and data transferred between them and the achieved throughput. The enhanced information given by Netflow comes with additional archival cost and many privacy concerns. To reduce the cost of collecting Netflow data, aggressive sampling (i.e., 1:1000 packets) is often employed, even for relatively low-speed networks [3]. Sampling affects the accuracy of Netflow data and may limit its applications [4]. Netflow records also include the IP addresses and port numbers used by the participating endpoints. Such content raises significant user privacy concerns [5], [6].

In this work, we propose leveraging SNMP link utilization data to accurately identify edge-to-edge (e2e) information about the achieved throughput of large network transfers. We have developed a methodology for inferring network events from SNMP traffic utilization time series data. Our method is the result of two main observations. First, looking at the time series of a link's utilization, we observe events where the utilization of the link increases (or decreases) to a different level, deviating from the link's normal behavior up to that point. These events could be considered as starting (or ending) points of high-throughput transfers. Second, these events propagate from the input links of a router to the output links of the same router, and from there to a neighboring router, allowing us to infer the actual route that the specific event followed.

Figure 1 illustrates these observations over a network example. Each router connects an organization's internal network to other organizations or intermediate routers. Using SNMP link usage data one can form the utilization time-series for each interface, which represents the traffic transferred between two connected routers. Looking at the time-series between $R7$ and $R9$ one can observe an increase in the link utilization. This increased utilization lasts for some time and then drops. Such behavior can be attributed to a transfer initiated from $R7$'s access network towards some destination. Following this increase from $R9$ to the next router and so on, we can observe that the corresponding transfer continues through $R12$ and $R14$. After $R14$ the transfer either continues to another network or is destined to a host in the access network served by $R14$. Note that the involved router interfaces do not have the same traffic variations in general. At the point that this transfer starts or ends, however, their traffic level changes in a similar fashion. Other transfers can be identified in different parts of the network at the same time. For example we can also observe a transfer between $R1 - R11$ and two transfers between $R2 - R6$.

The work presented in this paper is, to the best of our knowledge, the first that suggests the possibility of inferring edge-to-edge information from aggregated link utilization measurements. In a related work, Gerber et. al. used flow records to estimate the achievable download speed [7]. Similarly to our work, their algorithm eliminates the need for, network intrussive, active measurements. In contrary to this work, the use of flow records makes their solution expensive to deploy. Our contributions can be summarized as follows:

1) We propose a methodology to identify events in SNMP utilization time series.
2) We propose a methodology to map events in an input interface of a router to the output interface of the same router the event is switched to.

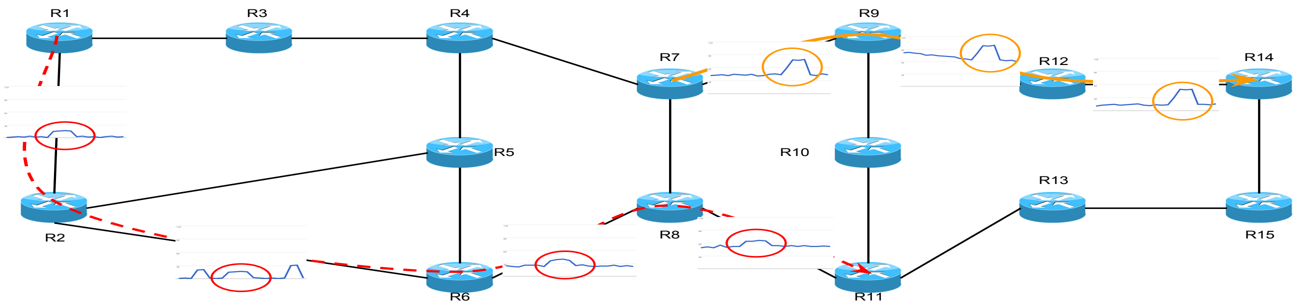


Figure 1. Several traffic utilization increase and decrease events can be identified in each observation period.

3) We evaluate our methodology and show that it can accurately identify events in a real network.

The rest of the paper is organized as follows. Section II provides a detailed description of our methodology also listing the specific challenges we came across in each step. In Section III we provide a detailed evaluation of the proposed methods. Section IV list a number of sample applications that can benefit from our method. Section V discusses open challenges for our method. Finally, Section VI concludes our work.

## II. Methodology

Out method consists of the following three steps:

**Event inference:** In the first step, we identify transfer-start and transfer-end events in the link utilization time series. This is an online processing step. We first transform the utilization time series into a 2-step differentiated time series, and identify as “events” those differences that are larger than a specific, user defined, threshold. The threshold can be defined based on the utilization and variations of the interface(s) of interest. We also propose a simple outlier detection method able to identify events by examining if the link’s utilization at the current time period (last 30 seconds) has deviated significantly from the utilization of the link in a recent time window.

**Mapping incoming events to outgoing interfaces:** After we identify an event (either transfer-start or transfer-end) at an input interface, we proceed to identify the output interface at the same router that the event is forwarded to. Our algorithm considers all transfer events that appear at any router input interface in that time period, and tries to find the most likely outgoing interface that each of those transfer events also appears in.

**Identify edge-to-edge path:** This step aims to identify the next router that each identified transfer is forwarded to. This step is accomplished easily when we have the network topology of the given network, including the IP address of every router interface in that network. If this information is not available it can be inferred by a matching process between the current output interface and all other input interfaces.

Note that all the three steps of our algorithm can be executed in real-time as new traffic utilization data become available for each link.

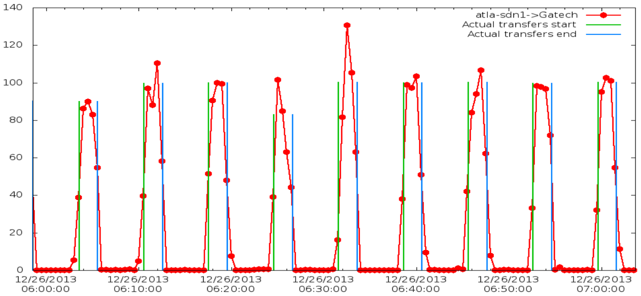


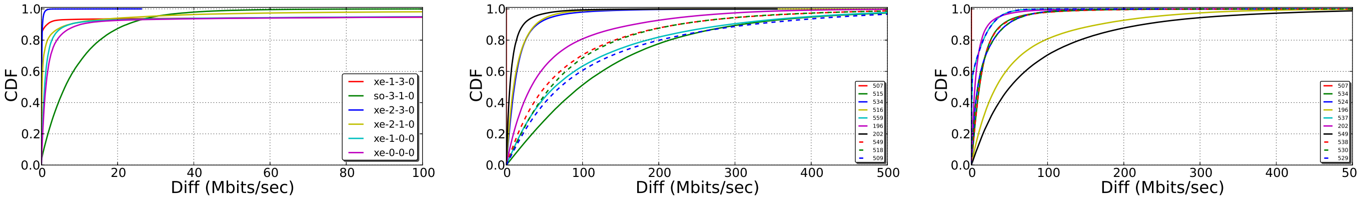
Figure 2. SNMP utilization time series of a link during a number of 100 Mbits/sec transfers traversing that link.

### A. Event inference

SNMP periodically, every $\Delta$ seconds, reports the number of bytes that traversed a specific router link over the previous interval $(t - \Delta)$. Using these byte counts, we can extract the average throughput utilization $U_i(t)$ for a link $i$ over the interval $(t - \Delta, t)$, $U_i(t) = \frac{Bytes(t-\Delta,t)\times 8}{\Delta}$ (1).

Using the link utilization time series we are interested in identifying changes in the utilization of a link $I$ that are created when a high-throughput flow starts or ends. We refer to these changes as flow events $e$ at interface $I$ and denote them with $I(e)$.

Figure 2 plots the SNMP traffic utilization time series as seen in a single network interface. A number of 100 Mbits/sec transfers were active during this time period, traversing the link. Vertical lines show the actual start (green) and end (blue) times of each transfer. We can observe the transfers to gradually appear in the SNMP utilization time series. This can be explained by two reasons: $(i)$ the actual flow events are not aligned with the utilization reporting times and $(ii)$ the utilization time series is an average over all the events at that $\Delta$. Depending on when the flow event appears relatively to the interval start it will affect the average differently. Considering these observations, just using the difference in the utilization between consecutive intervals gives misleading information regarding the flow’s throughput since it will only account for the difference in the interval $t$. If the flow started in the end of interval $t - 1$, then the difference will be close to the actual flow throughput. However, if the flow started towards the beginning or the middle of $t - 1$ then the difference will be far from the actual flow throughput. $V$ takes into account these non-alignment and averaging effects.



(a) ESnet router

(b) Commercial router: Busy links

(c) Commercial router: Highly variate links

Figure 3. CDF of event magnitude for different links and routers.

To identify the events we first transform the utilization time series of $I$ to the 2-step differential time series $V$, where $V_i(t) = U_i(t+1) - U_i(t-1)$ (2). $V$ provides the time series of the link utilization difference between two intervals with distance $2 \times \Delta$. The previous equation will result to the time-series of the utilization difference between two intervals with distance $2 \times \Delta$. This difference is more likely to be closer to the actual flow throughput since it allows for the transition of the traffic utilization from the base ($U_i(t-1)$ in this case) to base + flow throughput ($U_i(t+1)$), in the case of a flow start. In the following we describe two methods to identify these transitions.

*1) Threshold based event identification:* After this point we consider each value in the $V_i(t)$ time series as a possible transfer start (positive values) or end (negative values). We leave it up to the user to decide which of the events she considers significant. Figure 3 plots the CDF of the magnitude of all events for all the interfaces of an ESnet router located in Atlanta (left) and for the 10 most busiest (middle) and most variable (right) interfaces of an edge router connecting GaTech to the commercial Internet. We calculate variability for a link by estimating the Coefficient of Variation ($CoV = \frac{\sigma}{\mu}$) for the traffic time series of the interface. We then select the links with the highest CoV values for Figure 3(c). Note that the events' magnitude can vary significantly at different links. Some links allow for the identification of rather minor events, i.e., less than 3 Mbit/sec, while in other links most of the events are larger than 100 Mbits/sec. Depending on the link (or path) of interest the user can appropriately set the threshold value for event identification.

*2) Outlier based event identification:* The first step of this process is to identify significant changes in the differentiated time series $V$. To identify them we need an outlier detection method that is $(i)$ robust to the utilization variability, $(ii)$ robust to any periodicity in the time-series $(iii)$ does not assume any predefined distribution of the data and $(iv)$ is able to detect outliers online as new data become available. A simple such method is running a robust moving window average over the data, and estimating the Median Absolute Deviation (MAD) from the median during the observation window. This method is also known as the Hampel identifier [8], [9]. The method is controlled by two parameters $(a)$ the size $N$ of the observation window and $(b)$ the multiplicative factor $c$ that defines how strict the outlier detection method is. If $c$ is large then a value should be significantly far from the median to be identified as an outlier. If $c$ is small then values that create small deviations can also be considered outliers. We empirically examine appropriate values for these parameters in the next section.

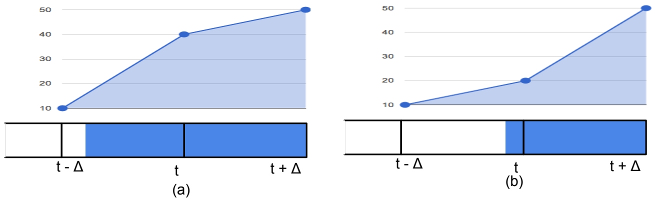


Figure 4. Event time approximation.

**Eliminating surrounding outliers:** As noticed earlier, a transfer event gradually appears in the utilization time series. As a result a single transfer event might correspond to more than one outliers. Each of the outliers corresponds to one step in the gradual appearance of the flow in the utilization time series. To accurately estimate the throughput of the event and also avoid having multiple values for each event we decide to keep only a single value for each sequence of consecutive outliers of the same trend. To do so, we sum, for each such sequence, the non-overlapping outliers, i.e., $V_i(t)$ for not consecutive $t$'s. For each outlier sequence we then keep the maximum of these sums.

**Event time approximation:** In both the above methods, the time $t$ of the identified event gives the time of the event relative to the closest utilization interval. To more accurately approximate the event time $T_e$ we use the following equation: $T_e = t - (\Delta \times \frac{|U(t)-U(t-\Delta)|}{|U(t+\Delta)-U(t-\Delta)|})$. Figure 4 illustrates how this approximation works. The plots show sample link utilization time series. The box underneath each plot shows a transfer starting at specific time. When the transfer started close to $t - \Delta$, it will count in the estimation of the throughput utilization $U_t$ in the interval $(t-\Delta, t)$ and thus result to a larger value for $U_{t-\Delta}$ (Figure 4(a)). This results to a large fraction, moving $T_e$ closer to the beginning of interval $t - \Delta$. In the opposite case that the transfer started closer to the end of the interval $(t-\Delta, t)$, the fraction will be small and the estimate $T_e$ will be closer to $t$ (Figure 4(b)).

### *B. Mapping input events to output interfaces*

The previous section described how we identify transfer events in the utilization time series of a link. Our next step is to identify the output interface the event will be forwarded to. This procedure is not trivial and may not have a definite answer at all times. That is because an identified event $E_i(t)$ at input $i$ may actually be the aggregation of a set of transfers $S(E_i(t))$, such as $E_i(t) = (e_x(t) \in S(E_i(t)))$, each with rate equal to $r(e_x(t))$, that appear at input $i$ at time $t$. The rate change $V_i(t)$

we observe associated with $E_i(t)$ is $V_i(t) = \sum r(e_x(t))$. The mapping problem then becomes: *Given a rate change $V_i(t)$ for every input and output i of a router, at a given time t, determine the switch mappings $I(e) \rightarrow O(e)$ for every transfer $e(t)$ at that time step.*

One can easily show that this problem cannot be solved in the general case. Consider, the case of two aggregated events $E_{I1}$ and $E_{I2}$ at input interfaces $I1$ and $I2$. Suppose that all transfers in those aggregate events are switched to two outputs, creating the aggregate events $E_{O1}$ and $E_{O2}$ at outputs $O1$ and $O2$ as follows: $E_{I1} = \{e_1, e_2\}$, $E_{I2} = \{e_3, e_4\}$, $E_{O1} = \{e_1, e_3\}$ and $E_{O2} = \{e_2, e_4\}$. The only relationship we can find in this scenario, considering all possible combinations of inputs or outputs, is $V_{I1} + V_{I2} = V_{O1} + V_{O2}$. This is obviously not sufficient to solve the problem defined above. To solve the problem we consider the following four conditions:

**Condition-1:** Every transfer appears individually at both its input and output interface. Assuming that each event has a distinct rate, we can solve the problem by identifying for every input rate change an (approximately) equal rate change at one of the router outputs.

**Condition-2:** Every event appears individually at its input but not necessarily so at its output. Assuming that any possible combination of events has a distinct aggregate rate, we need to find the set of inputs $I_j$ such that $\sum_{E_i \in I_j} V_{Ei} = O_j$, for every output interface $O_j$.

**Condition-3:** Every event appears individually at its output but not necessarily so at its inputs. Similarly to condition-2 we need to find the set of outputs $O_j$ such that $\sum_{E_i \in O_j} V_{Ei} = I_j$, for every input interface $I_j$.

**Condition-4:** Every event appears individually at its input or output or both. Without loss of generality, consider that an event $e$ appears individually at its input, say $I_e$. If this event appears individually also at its output, we can identify its switch mapping as in *Condition-1*. Otherwise, say that $e$ is part of an aggregate event $E_k$ at output $k$. Then, the rest of the events in $E_k$ must appear individually at their inputs (based on Condition-4). So, we can identify their switch mappings as we did in *Condition-2*. Similarly, if an event appears individually at its output.

Our algorithm for identifying the switch mappings first considers all possible combinations of output interfaces for every input interface to find a matching aggregate rate as in *Condition-3*. Then, it considers all possible combinations of inputs for every output to find a matching aggregate event, as in *Condition-2*. Since small rate variations may occur internally in the router, when traffic switches from the input interface to the output, we use the similarity function $S = \frac{||V_I| - |V_O||}{max(|V_I|, |V_O|)}$ (3), to compare the input and output variations at each step. We consider the mapping that minimizes $S$ as the most likely mapping between the input and output events of interest. An input combination might be also rejected if any of the individual interfaces in that combination (or smaller combinations) better matches the outgoing interface. Similarly for the reverse scenario. We evaluate appropriate similarity thresholds in the next section. Algorithm 1 presents the pseudocode of our switch mapping method.

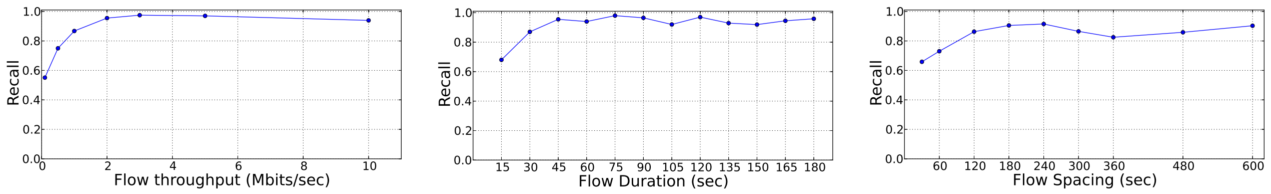
**Algorithm 1** Mapping Input to Output interfaces

```
procedure FIND_OUTPUT($E_i(t), E(t), V_o(t)$)
    $IN \leftarrow$ all combinations of $E_i(t)$ with events $E(t)$ from the other input interfaces
    $OUT \leftarrow$ all combinations of 2-step differential at time $t$ for all output interfaces $V_o(t)$
    $RES \leftarrow \emptyset$
    for all $o \in OUT$ do                          ▷ For all output event combinations
        $S = \frac{absE_i(t) - |o|}{max(E_i(t), o)}$          ▷ Calculate similarity with input event
        if $|S| \leq D$ then
            $RES \leftarrow |S|$
        end if
    end for
    for all $i \in OUT$ do                          ▷ For all input event combinations
        for all $e \in V_o(t)$ do                   ▷ for all single events in the output interfaces
            $S = \frac{|i| - |o|}{max(i, o)}$                ▷ Calculate their similarity
            if $|S| \leq D$ then
                $RES \leftarrow |S|$
            end if
        end for
    end for
    if $RES \neq \emptyset$ then
        $r \leftarrow$ argmin($RES$)               ▷ Get pair with the maximum similarity
        return $r$
    else
        return NULL
    end if
end procedure
```

### *C. Identify the edge-to-edge path*

This step aims to identify the edge-to-edge path the event will follow. In this paper we only consider the case where a complete connected view of the monitored network is available. That means that information for all routers the event traverses is available. If we already know which two routers a link connects, we can proceed from one router to the other. We can then infer the path by identifying the switch mappings for all routers in the path step by step, using the method described in the previous section. After we identify the outgoing interface in router $R_1$ we can then proceed to router $R_2$, that the corresponding link connects to. In this case the outgoing event $E_O(t)$ in interface $O$ of $R_1$ becomes the incoming event in the interface $I$ of $R_2$. Using the switch mapping method we can now identify the outgoing interface of the event in $R_2$.

In the case where we do not know which two end-points a link connects, we need to identify this hop using some of the available information. Since a link between two routers is a physical link both end points will most probably observe the exact traffic (with minimal variation due to reporting synchronization). With this fact in mind we can compare either the traffic utilization of the current output host with all input interfaces in the network and identify the incoming interface with the closest traffic utilization time series. One option is to use the Euclidean distance to calculate the distance ($d$) between each two interfaces for a time window $n$, ($d = \sqrt{\sum_{t=1}^{n} (U_o(t) - U_i(t))^2}$ (4)), An alternative approach would be to use the 2-step differentiated time series for calculating the distance, instead of the actual traffic utilization time series. Note that we do not need to run the above step for every identified event. Physical links do not change often and thus only verifying the routers connecting the links every few days should be enough.

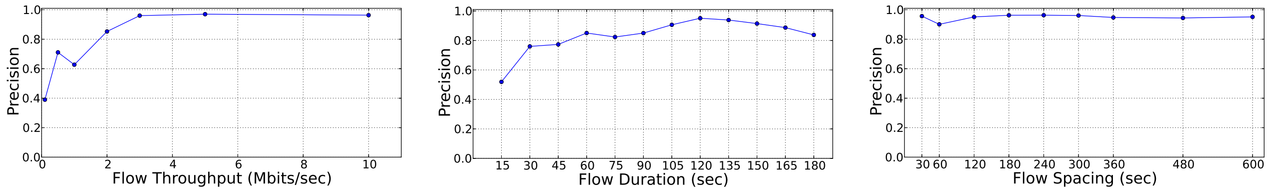


(a) Transfer throughput

(b) Transfer duration

(c) Transfer spacing

Figure 5. Recall when varying specific transfer properties ($c = 1$ and $W = 20min$).



(a) Transfer throughput

(b) Transfer duration

(c) Transfer spacing

Figure 6. Precision when varying specific transfer properties ($c = 1$ and $W = 20min$).

## III. Evaluation

### A. Datasets

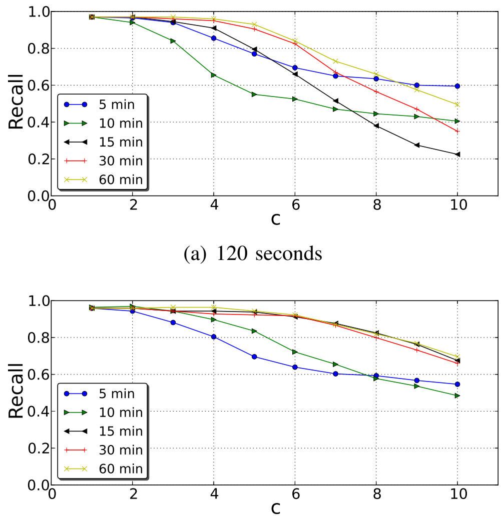
The methodology presented in the previous section aims at inferring the achieved throughput for transfers with a significant amount of data, i.e., high throughput and long duration transfers. In this section we try to identify the lower limits our method has in terms of $(i)$ *transfer throughput*, $(ii)$ *transfer duration* and $(iii)$ *spacing* between two consecutive transfers. Additionally, we identify appropriate values for the method parameters that allow for high accuracy transfer inference.

To evaluate our methodology we create a number of artificial datasets, composed by TCP transfers of specific characteristics, between a machine in Georgia Tech (GT) and Lawrence Berkeley National Lab (LBL). The created transfers traverse only the ESnet network, for which we have access to all the intermediate router utilization data. We also know the routers and interfaces each link connects. We use the nuttcp network performance tool to create the transfers. Depending on the desired transfer characteristic, we keep all other characteristics constant and vary the value of the characteristic in question.

We evaluate the accuracy of our algorithm by calculating $(i)$ recall and $(ii)$ precision. Recall is defined as the number of true positives $(TP)$, i.e., the number of actual transfer-events our method identified, over the total number of events we created. Precision is calculated as $\frac{TP}{TP+FP}$, where $FP$ is the number of false positives, i.e., events detected by our method that were not part of the artificial dataset.

### B. Outlier based event identification method

**Minimum event throughput:** Figure 5(a) plots recall as a function of the transfer's throughput. Our method manages to achieve more than 95% recall for transfers with throughput 2 Mbits/sec and larger. Figure 6(a) plots the precision of the method. Precision reaches values larger than 0.95 for transfer throughput larger than 3 Mbits/sec. We consider the latter value to be a reasonable throughput threshold.



(a) 120 seconds

(b) 180 seconds

Figure 7. Effect of TCP transfer duration in the selection of $w$ and $c$ values on method recall.

**Minimum transfer duration:** Figure 5(b) plots recall as a function of the transfer duration. We can observe that our method can achieve recall larger than 95% for duration larger than $\Delta$. Looking at the method's precision (Figure 6(b)), we can see that our method can achieve precision larger than 0.9 when the duration of the transfer is close to 2 minutes.

**Minimum transfer spacing:** Figure 5(c) plots recall as a function of the interval between the transfers. Recall increases to values larger than 85% when the transfer spacing is 2 minutes. After that point it stabilizes to similar or larger values. The precision remains high as the transfer spacing increases (Figure 6(c)).

**Method parameters:** Figure 7 examines how the choice of $W$ and $c$ affect the method's recall as transfer duration varies. Small window sizes only work well with small $c$ values,

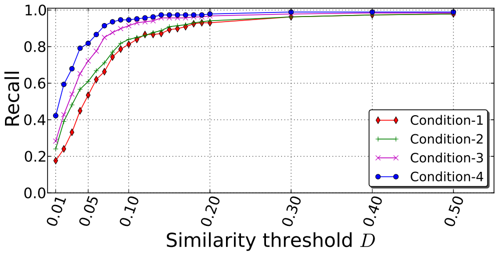


Figure 8. Recall as a function of the similarity threshold between input and output mappings.

resulting to low recall when $c > 2$. For both transfer duration values we can say that a window larger than 15 minutes and a $c$ smaller or equal to 5 provide acceptable accuracy. The precision did not show any variability with the method's parameters when $c > 1$ and $W > 5$ minutes.

### C. Traffic switch mapping

We now examine the accuracy of the method for inferring the outgoing interface an event, identified by the previous method, will be forwarded to. To evaluate this method we create a number of high throughput transfers ($>$ 3 Mbits/s). We use information from Paris traceroute as the ground truth for the interface mapping. A true positive (TP) in this case is an event where one of the possible options given by Paris traceroute was included in the outgoing mapping result. Figure 8 plots recall for all different conditions explained in Section II-B as a function of the similarity threshold $D$. We can observe that *condition-1* is the one that provides the mapping in most of the cases. Using *condition-4* we can get an improvement of about 15% in all cases. Regarding the similarity threshold, a difference of 10% ($D = 0.1$) gives recall values larger than 80% in all cases. *Condition-4* gives more than 95% recall values for any value of $D$ larger than 0.1.

### D. Edge-to-edge throughput inference

In the previous subsection, we focused on the evaluation of the event identification and switch mapping methods individually. In this subsection we evaluate the accuracy of our method in inferring edge-to-edge events and reporting their achieved throughput. To do so, we create a second dataset where we create number of transfers between GaTech and LBL. Using nuttcp we limit the average transfer throughput to values in the range of 5 - 110 Mbits/sec. Using our method, described in the previous section, we then try to identify these transfers in the SNMP utilization data.

Figure 9 plots the recall value as a function of the achieved throughput of each event. The left plot shows recall for the whole path between Gatech and LBL (only routers belonging to the ESnet infrastructure). We can see that our method achieves recall values larger than 0.5 for transfer events larger or equal to 20 Mbits/sec. For transfers larger than 200 Mbits/sec, recall ranges to values larger than 0.8. To understand the low recall values for small throughput values, Figure 9 plots the recall values for each router in the path. We can observe that in most routers we can achieve recall values larger than 0.8 independent of the transfer achieved throughput. For one router in the path the recall values drop to values close 0.5 for small throughput values and increase as the transfer throughput increases to values larger than 20 Mbits/sec. Our intuition behind this behavior is that the specific router is a busiest network hub, than the other routers. This means that traffic from different interfaces ("noise" traffic) in that router mixes with the traffic we created. This "noise" traffic affects the IN/OUT interface mapping of small events. To verify this intuition Figure 10 (a), plots the average traffic in the two interfaces of interest (input and output) during the time of the actual events we try to identify. Traffic in a specific instance is calculated as the sum of the traffic on all interfaces of interest in that instance. We can observe that recall drops as the traffic in these interfaces increase. Figure 10 (b) plots the recall as a function of the average traffic in all interfaces of the router. Looking at the average traffic in all the router interfaces we can see how busy the router is and how additional traffic from other interfaces affects out method. We can observe that as the traffic traversing the router increases, the recall drops. This suggests that in an overall busy router the mapping process becomes more difficult since additional traffic from other interfaces might merge with the traffic of interest.

## IV. Applicability of the Method

**Improvements of throughput prediction:** TCP throughput prediction applications are usually based on historical transfers between the end points of interest [10]. The extensive sampling and limited availability of NetFlow data usually limits the applicability of these type of prediction approaches. The achieved transfer throughput values inferred through our method can be used as additional samples in the presence of NetFlow data. Furthermore, in cases where NetFlow data are not available, our method can provide a sample of transfer throughput measurements.

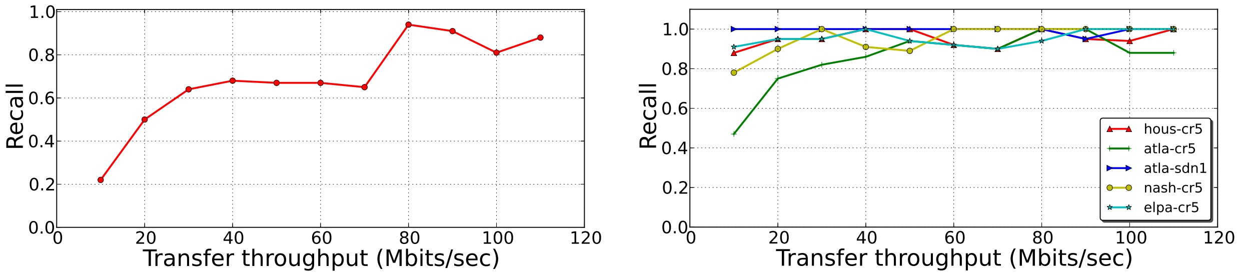
**DDoS attack initiator inference:** Spoofed traffic is a common method used by attackers to create Distributed Denial of Service (DDoS) attacks [11]. Using our method one can infer the actual source of spoofed traffic, by following the identified events to the source network and not relying to the IP address.

## V. Challenges

*Incomplete Data:* Access to SNMP data from every router in the network of interest is not always possible. For example, data may traverse routers that are not owned by the monitoring party, and thus cannot be collected. Additionally, equipment may fail and data might be lost for several reasons. We are exploring ways to take into account these cases, in our method, matching events that may appear to non-adjacent routers.

*Transfer identification:* An interesting next step would be to identify the actual transfers that traversed the path. This step would need to identify both transfer start and transfer end events and match them. This step is not trivial since transfer throughput might change through the duration of the transfer, or a number of transfers might be active during the same time at the path. We are considering several methods for transfer identification in our ongoing work.

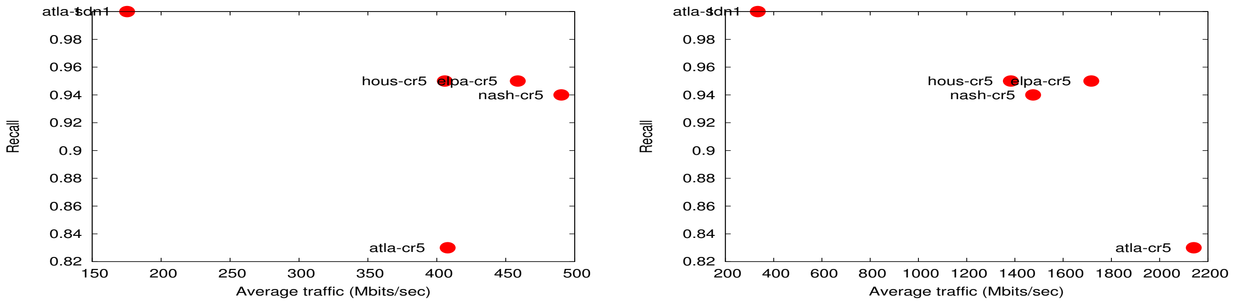
*Multipath transfers:* A common practice for load balancing network links is to use multiple paths to connect two networks. Some of the flows transferred simultaneously from a source



(a) Path recall vs. Actual transfer throughput

(b) Per router recall vs. Actual transfer throughput

Figure 9. Total path and per router Recall values as a function of the achieved transfer throughput.



(a) IN/OUT interfaces of interest

(b) All router interfaces

Figure 10. Per router in/out mapping recall vs. average traffic in router.

to a destination network are split among different paths that may not share any common routers apart from the source and destination ones. In this case, it is more difficult to identify a matching output link for an event observed in the source router, since split transfers will have lower magnitude. Currently, our methodology does not take these cases into account. We plan to further analyze the intermediate router variations in order to identify the several paths that the transfer could be traversed through. Also, correlation of non-adjacent router interfaces may give some indication for multipath transfers.

*Research Vs. Commercial Networks:* One may argue that ESnet, as a NREN, has very different traffic patterns than commercial networks and that our method would not apply in such data. Our results showed that we can also identify small magnitude events. In our future work we plan to investigate the applicability of our method in commercial environments with thousands of small flows starting and ending each measurement epoch.

## VI. Conclusion

In this paper, we provide evidence that using SNMP link counts we can infer the achieved throughput of Edge-to-Edge transfers taking place in the network. Our method first identifies significant variations in the link counts, and tags them as possible transfer starting or ending points. Iteratively following these variations to the neighboring routers, we are then able to identify the path the specific transfer traversed through the monitored network. Our ongoing work is designed to further evaluate our methodology. Additionally, we plan to test the applicability of the inferred e2e transfers to a number of applications, such as throughput prediction, traffic matrix estimation.

## VII. Acknowledgments

This research was supported by the U.S. Department of Energy under grant number DE-FG02-12ER26065.

## References

[1] J. Case, M. Fedor, M. Schoffstall, and J. Davin, “Simple network management protocol (snmp),” 1990.

[2] D. Harrington, R. Presuhn, and B. Wijnen, “An architecture for describing simple network management protocol (snmp) management frameworks,” rfc 3411, Dec, Tech. Rep., 2002.

[3] C. Estan and G. Varghese, “New directions in traffic measurement and accounting,” in ACM SIGCOMM CCR. ACM, 2002.

[4] B. Choi and S. Bhattacharyya, “On the accuracy and overhead of cisco sampled netflow,” in ACM LSNI, 2005.

[5] S. Coull, M. Collins, C. Wright, F. Monrose, and M. Reiter, “On web browsing privacy in anonymized netflows,” in Proceedings of 16th USENIX Security Symposium. USENIX Association, 2007.

[6] M. Foukarakis, D. Antoniades, S. Antonatos, and E. Markatos, “Flexible and high-performance anonymization of netflow records using anontool,” in SecureComm. IEEE, 2007.

[7] A. Gerber, J. Pang, O. Spatscheck, and S. Venkataraman, “Speed testing without speed tests: estimating achievable download speed from passive measurements,” in Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. ACM, 2010, pp. 424–430.

[8] L. Davies and U. Gather, “The identification of multiple outliers,” Journal of the American Statistical Association, vol. 88, no. 423, 1993, pp. 782–792. [Online]. Available: http://www.jstor.org/stable/2290763

[9] P. Menold, R. Pearson, and F. Allgower, “Online outlier detection and removal,” 1999.

[10] Q. He, C. Dovrolis, and M. Ammar, “On the predictability of large transfer tcp throughput,” in ACM SIGCOMM CCR. ACM, 2005.

[11] F. Lau, S. H. Rubin, M. H. Smith, and L. Trajkovic, “Distributed denial of service attacks,” in Systems, Man, and Cybernetics. IEEE, 2000.

# Slow Start TCP Improvements for Video Streaming Applications

Gaku Watanabe, Kazumi Kumazoe, Dirceu Cavendish, Daiki Nobayashi, Takeshi Ikenaga, Yuji Oie
Department of Computer Science and Electronics
Kyushu Institute of Technology
Fukuoka, Japan
e-mail: {i108132g@tobata.isc, kuma@ndrc, cavendish@ndrc, nova@ecs, ike@ecs, oie@ndrc}.kyutech.ac.jp

***Abstract*—Video streaming has become the major source of Internet traffic nowadays. In addition, content delivery network providers have adopted Video over HTTP/TCP as the preferred protocol stack for video streaming. In our previous work, we have shown how TCP variants play a definite role in the quality of video experience. In this paper, we research several mechanisms within TCP slow start phase to enhance video streaming experience. We utilize network performance measurers, as well as video quality metrics, to characterize the performance and interaction between network and application layers of video streams for various network scenarios. We show that video transport performance can be enhanced with small changes in TCP Slow Start.**

***Keywords*—*Video streaming; high speed networks; TCP congestion control; Packet retransmissions; Packet loss.***

## I. Introduction

Transmission control protocol (TCP) is the dominant transport protocol of the Internet, providing reliable data transmission for the large majority of applications. For data applications, the perceived quality of experience is the total transport time of a given file. For real time (streaming) applications, the perceived quality of experience involves not only the total transport time, but also the amount of data discarded at the client due to excessive transport delays. Transport delays depend on how TCP handles flow control and packet retransmissions. Therefore, video streaming user experience depends heavily on TCP performance. TCP protocol interacts with video application in non trivial ways. Widely used video codecs, such as H-264, use compression algorithms that result in variable bit rates along the playout time. In addition, TCP has to cope with variable network bandwidth along the transmission path. Network bandwidth variability is particularly wide over paths with wireless access links of today, where multiple transmission modes are used to maintain steady packet error rate under varying interference conditions. As the video playout rate and network bandwidth are independent, it is the task of the transport protocol to provide a timely delivery of video data so as to support a smooth playout experience.

In the last decade, many TCP variants have been proposed, mainly motivated by performance reasons. As TCP performance depends on network characteristics, and the Internet keeps evolving, TCP variants are likely to continue to be proposed. Most of the proposals deal with congestion window size adjustment mechanism, which is called congestion avoidance phase of TCP, since congestion window size controls the amount of data injected into the network at a given time. In prior work, we have introduced a delay based TCP window flow control mechanism that uses path capacity and storage estimation [5] [6]. The idea is to estimate bottleneck capacity and path storage space, and regulate the congestion window size using a control theoretical approach. Two versions of this mechanism were proposed: one using a proportional controlling equation [5], and another using a proportional plus derivative controller [6]. More recently, we have studied TCP performance of most popular TCP variants - Reno [2], Cubic (Linux) [13], Compound (Windows) [14] - as well as our proposed TCP variants: Capacity and Congestion Probing (CCP) [5], and Capacity Congestion Plus Derivative (CCPD) [6], in transmitting video streaming data over wireless path conditions. Our proposed CCP and CCPD TCP variants utilize delay based congestion control mechanism, and hence are resistant to random packet losses experienced in wireless links.

In this paper, we show that it is possible to improve Slow Start mechanism of TCP to improve video streaming over Internet paths with wireless access links. More specifically, we demonstrate that: i) Open loop nature of slow start negatively affects video rendering quality; ii) Dampening congestion window growth during slow start may negatively affect video streaming performance; iii) Shortening slow start phase may improve video performance. The material is organized as follows. Related work discussion is provided on Section II. Section III describes video streaming over TCP system. Section IV introduces the TCP variants addressed in this paper, as well as additional mechanisms to enhance video streaming experience. Section VI addresses video delivery performance evaluation for each TCP variant and attempted enhancements. Section VII addresses directions we are pursuing as follow up to this work.

## II. Related Work

The impact of wide variability of TCP throughput caused by network packet losses on video streaming has been addressed [10] [4]. In [10], variable rate video encoders are considered, where video source adjusts its encoding rate according with network available bandwidth in the streaming path. In [4], a TCP Reno delay model is used by the video encoder to change encoding mode according with network conditions. Both approaches require a tight coupling between application and transport protocol. These approaches are opposite to what is taken in this work, which seeks to adjust video transport to arbitrary video encoders.

Modifications of TCP protocol to enhance video streaming have been recently proposed. Pu et al. [12] have proposed a proxy TCP architecture for higher performance on paths with last hop wireless links. The proxy TCP node implements a variation of TCP for which additive increase multiplicative decrease (AIMD) congestion window $cwnd$ adjustment is disabled, and replaced with a fair scheduler at the entrance of the wireless link. The approach, however, does not touch TCP sender at the video server side, which limits overall video streaming performance as characterized in [8].

A different approach to improving the transport of video streams is presented by [11]. Their work seeks to improve video streaming performance by streaming over multiple paths, as well as adapting video transmission rates to the network bandwidth available. Such approach, best suited to distributed content delivery systems, requires coordination between multiple distribution sites. In contrast, we seek to improve each network transport session carrying a video session by adapting TCP source behavior, independently of the video encoder.

Another distinct aspect of our current work is that we propose improvements on Slow Start, which is widely used by many TCP variants, on real client and server network stacks that are widely deployed for video streaming, via VLC open source video client, and standard HTTP server. We seek to understand how small changes in TCP Slow Start may affect the quality of user experience.

## III. Video Streaming over TCP

Video streaming over HTTP/TCP involves an HTTP server side, where video files are made available for streaming upon HTTP requests, and a video client, which places HTTP requests to the server over the Internet, for video streaming. Fig. 1 illustrates video streaming components.

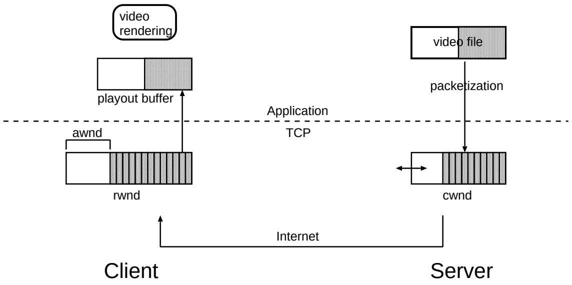


Fig. 1: Video Streaming over TCP

An HTTP server stores encoded video files, available upon HTTP requests. Once a request is placed, a TCP sender is instantiated to transmit packetized data to the client machine. At TCP transport layer, a congestion window is used for flow controlling the amount of data injected into the network. The size of the congestion window, $cwnd$, is adjusted dynamically, according to the level of congestion in the network, as well as the space available for data storage, $awnd$, at the TCP client receiver buffer. Congestion window space is freed only when data packets are acknowledged by the receiver, so that lost packets are retransmitted by the TCP layer. At the client side, in addition to acknowledging arriving packets, TCP receiver sends back its current available space $awnd$, so that at the sender side, $cwnd \leq awnd$ at all times. At the client application layer, a video player extracts data from a playout buffer, filled with packets delivered by TCP receiver from its buffer. The playout buffer is used to smooth out variable data arrival rate.

### A. Interaction between Video streaming and TCP

At the server side, HTTP server retrieves data into the TCP sender buffer according with $cwnd$ size. Hence, the injection rate of video data into the TCP buffer is different than the video variable encoding rate. In addition, TCP throughput performance is affected by the round trip time of the TCP session. This is a direct consequence of the congestion window mechanism of TCP, where only up to a $cwnd$ worth of bytes can be delivered without acknowledgements. Hence, for a fixed $cwnd$ size, from the sending of the first packet until the first acknowledgement arrives, a TCP session throughput is capped at $cwnd/rtt$. For each TCP Slow Start variant, to be described shortly, the size of the congestion window is computed by a specific algorithm at time of packet acknowledgement reception by the TCP source. However, for all TCP variants, the size of the congestion window is capped by the available TCP receiver space $awnd$ sent back from the TCP client.

At the client side, the video data is retrieved by the video player into a playout buffer, and delivered to the video renderer. Playout buffer may underflow, if TCP receiver window empties out. On the other hand, playout buffer overflow does not occur, since the player will not pull more data into the playout buffer than it can handle.

In summary, video data packets are injected into the network only if space is available at the TCP congestion window. Arriving packets at the client are stored at the TCP receiver buffer, and extracted by the video playout client at the video nominal playout rate.

## IV. Anatomy of transmission control protocol

TCP protocols fall into two categories, delay and loss based. Advanced loss based TCP protocols use packet loss as primary congestion indication signal, performing window regulation as $cwnd_k = f(cwnd_{k-1})$, being ack reception paced. Most $f$ functions follow an Additive Increase Multiplicative Decrease strategy, with various increase and decrease parameters. TCP NewReno and Cubic are examples of AIMD strategies. Delay based TCP protocols, on the other hand, use queue delay information as the congestion indication signal, increasing/decreasing the window if the delay is small/large, respectively. Vegas, CCP and CCPD are examples of delay based protocols.

Most TCP variants follow TCP Reno phase framework: slow start, congestion avoidance, fast retransmit, and fast recovery.

- **Slow Start(SS) :** This is the initial phase of a TCP session, where no information about the session path is assumed. In this phase, for each acknowledgement received, two more packets are allowed into the network. Hence, congestion window $cwnd$ is roughly doubled at each round trip time. Notice that the $cwnd$ size can only increase in this phase. Therefore, in this phase there is no flow control of the traffic into the network. This phase ends when the $cwnd$ size reaches a large value, dictated by $ssthresh$ parameter, or when the first packet loss is detected, whichever comes first. All widely used TCP variants make use of the same slow start except Cubic [13].

- **Congestion Avoidance(CA) :** This phase is entered when the TCP sender detects a packet loss, or the $cwnd$ size reaches a target upper size called $ssthresh$ (slow start threshold). The sender controls the $cwnd$ size to avoid path congestion. Each TCP variant has a different method of $cwnd$ size adjustment.
- **Fast Retransmit and fast recovery(FR) :** The purpose of this phase is to freeze all $cwnd$ size adjustments in order to take care of retransmissions of lost packets.

Fig. 2 illustrates various phases of a TCP session. Our interest is in the Slow Start phase of TCP, at the left of the figure, which dictates how much traffic is allowed into the network before congestion avoidance takes place. A comprehensive tutorial of TCP features can be found in [1].

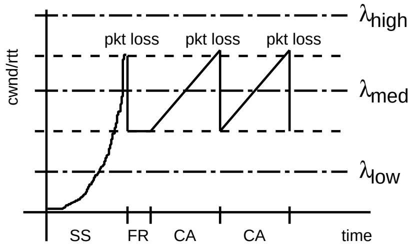


Fig. 2: TCP Congestion Window Dynamics vs Video Playout

For most TCP variants widely used today, with exception of Cubic, the slow start phase may negatively impact video experience, for two reasons: i) the amount of traffic injected into the network path is independent of the network path conditions, as well as the video playout rate; ii) Slow start phase typically ends with a large $cwnd$ size, which causes multiple packet losses, and may trigger long fast retransmit/recovery periods. During these periods, no new packets are admitted into the network until all lost packets are successfully delivered. For video streams, many of these lost packets may arrive too late for frame rendering, causing excessive frame discards. Finally, a responsive congestion avoidance mechanism affords quick adaptation to network conditions, which decreases playout buffer underflows as well as picture discards.

In this paper, we use CCP as a framework upon which we design Slow Start variation mechanisms. TCP CCP was our first attempt to design a delay based congestion avoidance scheme based on solid control theoretical approach. The cwnd size is adjusted according to a proportional controller control law. The cwnd adjustment scheme is called at every acknowledgement reception, and may result in either window increase and decrease. In addition, packet loss does not trigger any special cwnd adjustment. CCP cwnd adjustment scheme is as per 1:

$$cwnd_k = \frac{[Kp(B - x_k) - in\_flight\_segs_k]}{2} \quad 0 \leq Kp \qquad (1)$$

where $Kp$ is a proportional gain, $B$ is an estimated storage capacity of the TCP session path, or virtual buffer size, $x_k$ is the level of occupancy of the virtual buffer, or estimated packet backlog, and $in\_flight\_segs$ is the number of segments in flight (unacknowledged). Typically, CCP cwnd dynamics exhibit a dampened oscillation towards a given cwnd size, upon cross traffic activity. Notice that $cwnd_k$ does not depend on previous cwnd sizes, as with the other TCP variants. This fact guarantees a fast responsiveness to network bandwidth variations.

As CCP uses the same Slow Start mechanism as of most TCP variants widely deployed nowadays, all lessons learn on this paper are applicable to other TCP variants.

Let $\lambda$ be the video average bit rate across its entire playout time. That is, $\lambda = VideoSize/TotalPlayoutTime$. Fig. 2 illustrates three video playout rate cases: $\lambda_{high}, \lambda_{med}, \lambda_{low}$:

$\lambda_{high}$ The average playout rate is higher than the transmission rate. In this case, playout buffer is likely to empty out, causing buffer underflow condition.

$\lambda_{med}$ The average playout rate is close to the average transmission rate. In this case, buffer underflow is not likely to occur, affording a smooth video rendering at the client.

$\lambda_{low}$ The average playout rate is lower than the transmission rate. In this case, playout buffer may overflow, causing picture discards due to overflow condition. In practice, this case does not happen if video client pulls data from the TCP socket, as it is commonly the case. In addition, TCP receiver buffer will not overflow either, because $cwnd$ at the sender side is capped by the available TCP receiver buffer space $awnd$ reported by the receiver.

## V. TCP Slow Start Improvements for Video Streaming

We focus on the slow start phase, since as mentioned before, it is a phase at which much harm can be done to video streaming due to its open feedback nature. The original idea of Slow Start was to increase $cwnd$ as quickly as possible to large values, so that more data throughput could be achieved on a short period of time. However, for video applications, the ideal throughput should be close to the video rendering rate. So, there is no use in targeting too high throughput. For these changes, we target our CCP TCP variant, since we have control over its implementation.

- **ShortSlowStart:** In this scheme, our TCP variant (CCPSSS) in slow start attempts to transition into congestion avoidance as quickly as possible. However, because CCP requires estimation of network path capacity in congestion avoidance phase, the scheme waits until a first capacity estimate is available to transition out of slow start.
- **VideoRateStart:** In this scheme, TCP variant (CCPVS) in slow start attempts to set its $cwnd$ to a size such that its resulting throughput ($cwnd/rtt$) be close to the video average playout rate. Hence, $cwnd$ is expected to stabilize on a given value, during Slow Start.
- **VideoRateSsthresh:** In this scheme (CCPLSS), $ssthresh$ parameter, which caps the maximum value of the $cwnd$ before TCP leaves slow start and transitions

into congestion avoidance, is adjusted to correspond to the average video playout rate. That is, the scheme delivers a throughput around the video playout rate at the time of transition out of slow start.

## VI. Video Streaming Performance Characterization for various Slow Starts

Fig. 3 describes the network testbed used for emulating a network path with wireless access link. An HTTP video server and a VLC client machine are connected to two access switches, which are connected to a link emulator, used to adjust path delay and inject controlled random packet loss. All links are 1Gbps, ensuring plenty of network capacity for many video streams between client and server. No cross traffic is considered, as this would make it difficult to isolate the impact of TCP Slow Start mechanisms on video streaming performance.

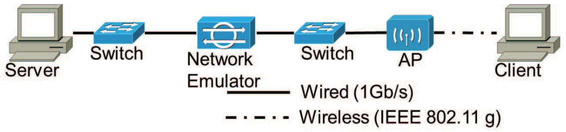


Fig. 3: Video Streaming Emulation Network

Video and network settings are as follows: video file size: $409Mbytes$; Playback time: $10min24sec$; Average playback rate: $5.24Mbps$; Encoding: MPEG-4; video codec: H.264/AVC; frame rate: 30fps; audio codec: MPEG-4 AAC; playout buffer size: $656Kbytes$, which drives initial buffering delay. TCP sender and receiver maximum buffer size: $256Mbytes$. The VLC client is attached to the network via a WiFi link. Iperf is used to measure the available wireless link bandwidth, to make sure it is higher than the average video playout rate. Packet loss is hence induced only by the wireless link, and is reflected in the number of TCP packet retransmissions.

Performance measurers adopted, in order of priority, are:

- **Picture discards:** number of frames discarded by the video decoder. This measurer defines the number of frames skipped by the video rendered at the client side.
- **Buffer underflow:** number of buffer underflow events at video client buffer. This measurer defines the number of "catch up" events, where the video freezes and then resumes at a faster rate until all late frames have been played out.
- **Packet retransmissions:** number of packets retransmitted by TCP. This is a measure of how efficient the TCP variant is in transporting the video stream data. It is likely to impact video quality in large round trip time path conditions, where a single retransmission doubles network latency of packet data from an application perspective.

We organize our test cases into the following categories:

- Typical round trip time
- Short round trip time
- Large round trip time

For each of these categories, we have run five trial experiments for each TCP Slow Start variant. Results are reported as average and standard deviation bars.

### A. Typical round trip time

Fig. 4 zooms into the first 30 secs of video streaming, to highlight $cwnd$ dynamics during slow start (X-axis in units of 100msecs). CCP(1) shows a standard $cwnd$ ramp up (first 2 secs of session), where $cwnd$ size is doubled at every round trip time. CCPLSS also shows a similar $cwnd$ ramp up, which is because the only difference is the $ssthresh$ value used. In contrast, CCPVS and CCPSSS rapidly ramp up their $cwnd$ size to a high value. CCPSSS shows oscillations on $cwnd$ size due to early poor capacity estimation.

Fig. 5 reports on bottleneck capacity estimation of each TCP variant. The graphs show that there is no difference in capacity estimation across the TCP variants. This is not surprising, given that the capacity estimation method, based on a packet pair dispersion measurement, is the same across all variants. Each TCP session sees around 15 Mbps bottleneck capacity, which in our experiment topology corresponds to the available bandwidth of the wireless WiFi access link.

Fig. 6 reports on the average packet round trip time seen by each TCP variant along the video streaming session. The graphs' dynamics has a 100msec support line, which is the round trip propagation delay between video source and client. In addition, the higher and more dense these graphs are, the more aggressive the TCP variant is. All variants present a highly dynamic rtt variation, which poses a challenge in video rendering from the video playout buffer.

Fig. 7 depicts video streaming and TCP performance under typical propagation delay of 100msecs. In this case, CCP(1) delivers best video experience, followed by CCPLSS. All variants feature small to negligible packet retransmissions except CCP(1), which is evidence of how vanilla Slow Start TCP hurts video streaming.

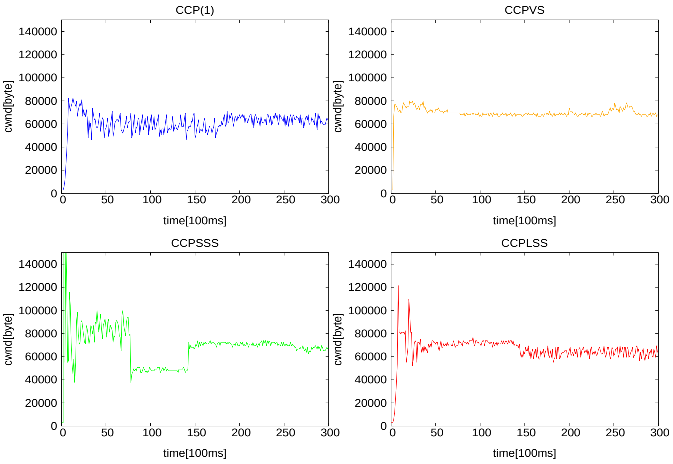


Fig. 4: Congestion window: first 30 secs; rtt=100msec

### B. Short round trip time

In this experimental settings, VLC client and server are connected via a very short path, with propagation delay of

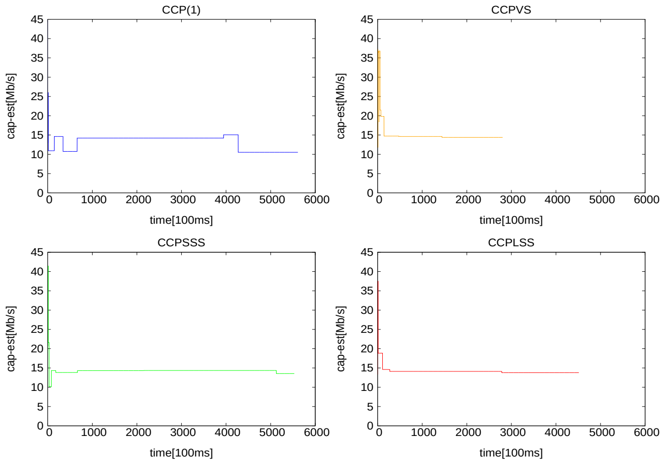


Fig. 5: Capacity estimation; rtt=100msec

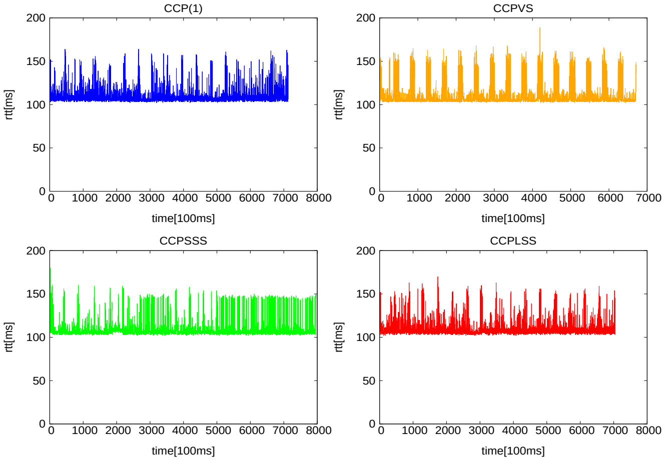
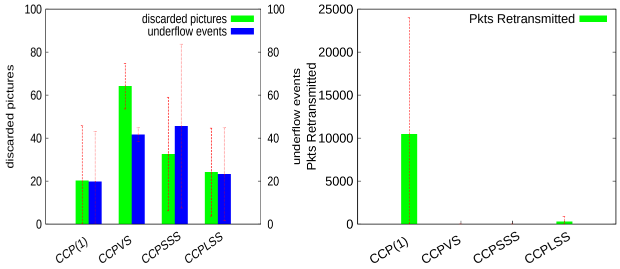


Fig. 6: Packet delay; rtt=100msec

around 3msec. Fig. 8 depicts the *cwnd* evolution of the first 30 secs of a video experiment trial. For all variants, we can see various Slow Start periods, where *cwnd* returns to small values and ramps up again. This is because the network path has very small space for packet storage, hence most congestion may be considered severe, which causes the TCP session to return to Slow Start phase.

Fig. 9 depicts packet round trip times measured between TCP sender and receiver. We first notice that this measurement includes retransmission at the network and WiFi link layers. The much inflated round trip time values measured as compared with propagation delay of around 3msec points to a large number of packet retransmissions, which is expected under short network path storage space. CCP(1) is clearly the most aggressive TCP variant, incurring in more packet delays than the other variants. CCPLSS seems to present a compromise between low and high round trip times. Notice that a constantly low rtt is not necessarily good, as it may indicate that the TCP variant is not responsive to network load and bottleneck variations throughout the streaming session.

Fig. 10 shows the video and TCP performance for this short rtt propagation delay scenario. In this case, CCPLSS delivers best video streaming quality, with least number of playout buffer underflows, as well as the least number of discarded pictures. Performance of all TCP variants are sharply different because of the repeated return of TCP session to Slow Start, due to the short network path (small storage space).



a) VLC performance b) TCP packets retransmitted

Fig. 7: Video Performance vs TCP performance; rtt=100msec

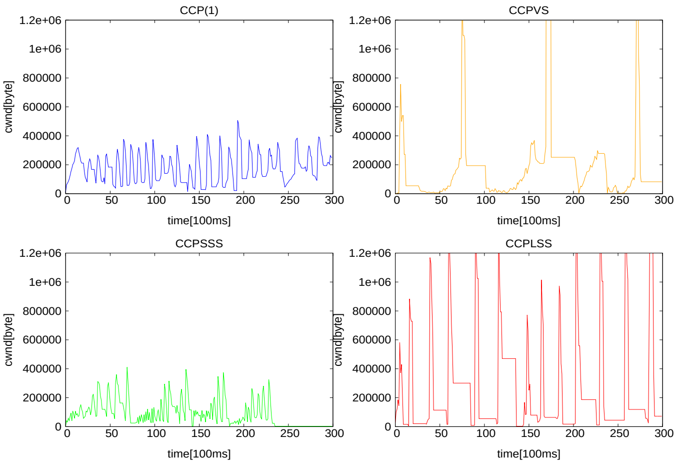


Fig. 8: Congestion window: first 30 secs; rtt=3msec

### C. Large round trip time

Fig. 11 zooms into the first 30 secs of video streaming, for the case of long round trip propagation delay. Notice that the *cwnd* oscillations are larger, which is expected for feedback loops with large dead time (propagation delay). Despite large *cwnd* oscillations, packet delays, not shown for space considerations, are more flat, due to more storage capacity in the long network path.

Finally, Fig. 12 reports VLC and TCP performance for long propagation delay. This long path case confirms that for high path storage conditions CCP(1) delivers best video streaming performance. This is evidence that aggressive TCP protocols deliver better video streaming performance when network path has enough storage space.

In our performance evaluation, we have not attempted to tune VLC client to minimize frame discards, even though VLC settings may be used to lower the number of frame discards. In addition, no tuning of TCP parameters was performed to better video client performance for any of the TCP variants studied. We have simply used parameter values from our previous study of CCP performance of file transfers [7]. All changes were limited to the Slow Start phase of TCP.

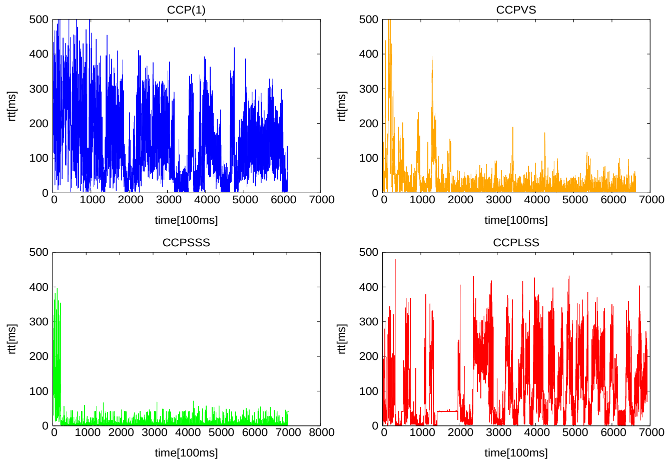


Fig. 9: Packet delay; rtt=3msec

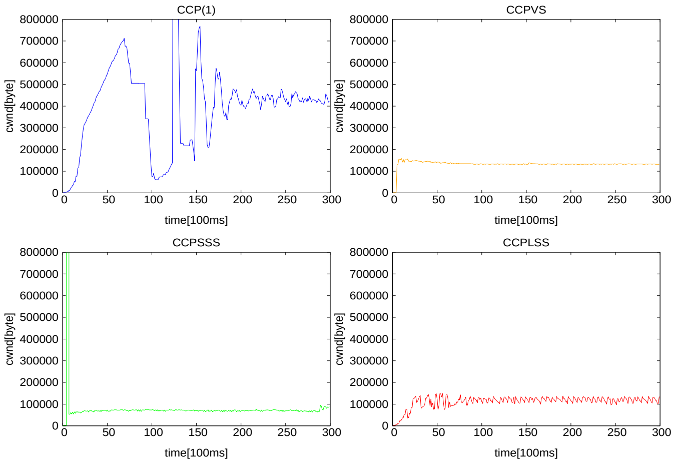
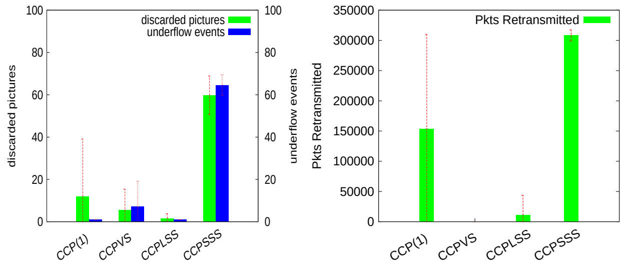
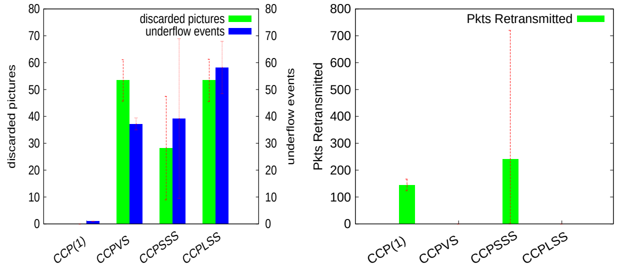


Fig. 11: Congestion window: first 30 secs; rtt=200msec



a) VLC performance b) TCP packets retransmitted

Fig. 10: Video Performance vs TCP performance; rtt=3msec



a) VLC performance b) TCP packets retransmitted

Fig. 12: Video Performance vs TCP performance; rtt=200msec

## VII. Conclusion and Future work

In this paper, we have introduced and evaluated a few variations of the slow start phase of TCP protocol to improve TCP transport performance of video streams. We have characterized TCP performance with these schemes when transporting video streaming applications over wireless network paths via open source experiments. Our experimental results show that tuning $ssthresh$ parameter to deliver video playout rate throughput when TCP transitions out of Slow Start delivers better Video Streaming performance on short network paths. As future work, we are currently exploring changes in the congestion avoidance phase of TCP, in order to further improve video streaming.

## Acknowledgments

Work supported in part by JSPS Grant-in-Aid for Scientific Research (B) (No 23300028).

## References

[1] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, “Host-to-Host Congestion Control for TCP, ” IEEE Communications Surveys & Tutorials, Third Quarter 2010, Vol. 12, No. 3, pp. 304-342.

[2] M. Allman, V. Paxson, and W. Stevens, “TCP Congestion Control,” IETF RFC 2581, April 1999.

[3] A. Ahmed, S.M.H. Zaidi, and N. Ahmed, “Performance evaluation of Transmission Control Protocol in mobile ad hoc networks, ” IEEE Int. Networking and Communication Conference, June 2004, pp. 13-18.

[4] A. Argyriou, “Using Rate-Distortion Metrics for Real-Time Internet Video Streaming with TCP, ” IEEE ICME06, 2006, pp. 1517-1520.

[5] D. Cavendish, K. Kumazoe, M. Tsuru, Y. Oie, and M. Gerla, “Capacity and Congestion Probing: TCP Congestion Avoidance via Path Capacity and Storage Estimation,” IEEE Second International Conference on Evolving Internet, best paper award, September 2010, pp. 42-48.

[6] D. Cavendish, H. Kuwahara, K. Kumazoe, M. Tsuru, and Y. Oie, “TCP Congestion Avoidance using Proportional plus Derivative Control,” IARIA Third International Conference on Evolving Internet, best paper award, June 2011, pp. 20-25.

[7] D. Cavendish et al., “On Tuning TCP for Superior Performance on High Speed Path Scenarios,” IARIA Fourth International Conference on Evolving Internet, best paper award, June 2012, pp. 11-16.

[8] G. Watanabe, K. Kumazoe, D. Cavendish, D. Nobayashi, T. Ikenaga, and Y. Oie, “Performance Characterization of Streaming Video over TCP Variants,” IARIA Fifth International Conference on Evolving Internet, best paper award, June 2013, pp. 16-21.

[9] S. Henna,“A Throughput Analysis of TCP Variants in Mobile Wireless Networks,” Third Int. Conference on Next Generation Mobile Applications, Services and Technologies - NGMAST, Sept. 2009, pp.279-284.

[10] P. Papadimitriou, “An Integrated Smooth Transmission Control and Temporal Scaling Scheme for MPEG-4 Streaming Video,” In Proceedings of IEEE ICME 08, 2008, pp. 33-36.

[11] J-W. Park, R. P. Karrer, and J. Kim,, “TCP-RomeL A Transport-Layer Parallel Streaming Protocol for Real-Time Online Multimedia Environments,” In Journal of Communications and Networks, Vol.13, No. 3, June 2011, pp. 277-285.

[12] W. Pu, Z. Zou, and C. W. Chen, “New TCP Video Streaming Proxy Design for Last-Hop Wireless Networks,” In Proceedings of IEEE ICIP 11, 2011, pp. 2225-2228.

[13] I. Rhee, L. Xu, and S. Ha, “CUBIC for Fast Long-Distance Networks,” Internet Draft, draft-rhee-tcpm-ctcp-02, August 2008.

[14] M. Sridharan, K. Tan, D. Bansal, and D. Thaler, “Compound TCP: A New Congestion Control for High-Speed and Long Distance Networks,” Internet Draft, draft-sridharan-tcpm-ctcp-02, November 2008.

[15] S. Waghmare, A. Parab, P. Nikose, and S.J. Bhosale, “Comparative analysis of different TCP variants in a wireless environment,” IEEE 3rd Int. Conference on Electronics Computer Technology, April 2011, Vol.4, pp.158-162.

# Purpose-bound Certificate Enrollment in Automation Environments

Rainer Falk and Steffen Fries
Corporate Technology
Siemens AG
Munich, Germany
e-mail: {rainer.falk|steffen.fries}@siemens.com

***Abstract*—Information security is gaining increasing importance for networked control systems. Examples are industrial automation, process automation, and energy automation systems. Characteristic for all these systems is the data exchange between intelligent electronic devices – IEDs, which are used to monitor and control the operation. In energy automation these IEDs provide the data for a obtaining a system view of connected decentralized energy resources – DER. Based on the system view, a set of DER building a virtual power plant (VPP) can be managed reliably. The communication is realized through domain-specific protocols like IEC 61850 or IEC 60870-5. The communication is performed increasingly also over public networks. Therefore IT security is a necessary prerequisite to prevent intentional manipulations, thereby ensuring the reliable operation of the energy grid. Basis for protecting metering and control communication are cryptographic security credentials, which need to be managed not only during operation, but most importantly during installation (initial enrollment). This process needs to be as simple as possible to not increase the overall effort and to not introduce additional sources for failures. Hence, automatic credential management is needed to ensure an efficient management for a huge number of devices. This paper describes a new approach for the automatic initial security credential enrollment process during the installation phase of IEDs. The approach targets the binding of the installed IEDs to the operational environment and also to the intended utilization of the IED by embedding specific information into the enrollment communication, which is then reflected in the issued X.509 certificates.**

***Keywords–security; device authentication; certificate enrollment; real-time; network access authentication; firewall; substation automation; smart grid; IEC 61850, IEC 60870-5, IEC 62351***

## I. Introduction

Decentralized energy generation, e.g., through renewable energy sources like solar cells or wind power, is becoming increasingly important to generate environmentally sustainable energy and thus to reduce greenhouse gases leading to global warming. Introducing decentralized energy generators into the current energy distribution network poses great challenges for energy automation as decentralized energy generation needs to be monitored and controlled to a similar level as centralized energy generation in power plants. This requires widely distributed communication networks. Distributed energy generators may also be aggregated on a higher level to form a so-called virtual power plant. Such a virtual power plant may be viewed from the outside in a similar way as a common power plant with respect to energy generation capacity. But due to its decentralized nature, the demands on communication necessary to control the virtual power plant are much more challenging. Moreover, these decentralized energy resources may also be used in an autonomous island mode, without any connection to a backend system.

Furthermore, the introduction of controllable loads on residential level requires enhancements to the energy automation communication infrastructure as used today. Clearly, secure communication between a control station and equipment of users (e.g., decentralized energy generators) as well as with decentralized field equipment must be addressed. Standard communication technologies as Ethernet and the Internet protocol IP are increasingly used in energy automation environments down to the field level [1] [2].

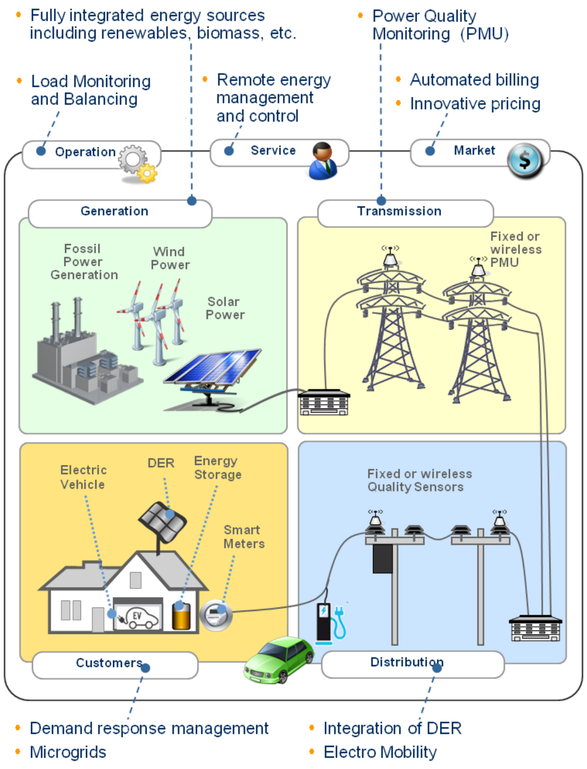


Figure 1. Typical Smart Grid Scenarios

Figure 1 depicts example Smart Grid scenarios showing the increased communication demand, e.g., through the integration of microgrids, controllable loads, and also electro mobility. IT security is a base requirement to be addressed in all the scenarios to ensure the reliable operation of the smart grid. One base for the secure interaction are typically security credentials in the form of X.509 digital certificates, corresponding private keys, and a related security policy. All need to be provisioned during device installation and maintained during operation. Especially the exchange of devices with spare parts should not lead to breaches in security, which could occur if the key material of the replaced devices is not handled appropriately. To ensure that this key material cannot be misused, e.g., in the context of an unintended service or in an unintended environment, the key material has to be bound to the respective device purpose. Existing options, e.g., using key usage extensions in X.509 certificates, may not always be sufficient, as they relate to the actual usage of the cryptographic key and not to the device application environment. The purpose-binding of a cryptographic key described in this paper therefore restricts the key acceptance depending on location information, and potential other parameters.

The remainder of this paper is structured as follows: Section II provides an overview of two example Smart Grid use cases. Section III depicts an overview of secure communication with respect to the use cases explained before. This section motivates the handling of security key material. Section IV introduces the Public Key Infrastructure as means for credential handling. Section VI introduces existing certificate enrollment methods, while Section VI describes an enhancement to have purpose bound certificates. Section VII concludes the paper and provides an outlook.

## II. SMART GRID USE CASES

To motivate communication security, two example use cases are addressed in this paper, substation automation and DER incorporation in energy control networks. They are explained in the following two subsections.

### A. *Substation Automation*

Automation networks are typically shared networks connected in a ring, star, or bus topology, or a mixture of these. Most often, the time-critical part is realized on a dedicated network segment, while the rest of the communication supporting the automation systems is performed on networks with lower performance requirements.

An example for energy automation is the communication within a substation. A substation typically transforms voltage levels, and includes power monitoring and protection functions. The example in Figure 2 shows a typical setup of a primary substation. The red rectangle shows the area, in which the IEDs communicate status information and provide this information into the substation automation zone and further up the hierarchy to the control center.

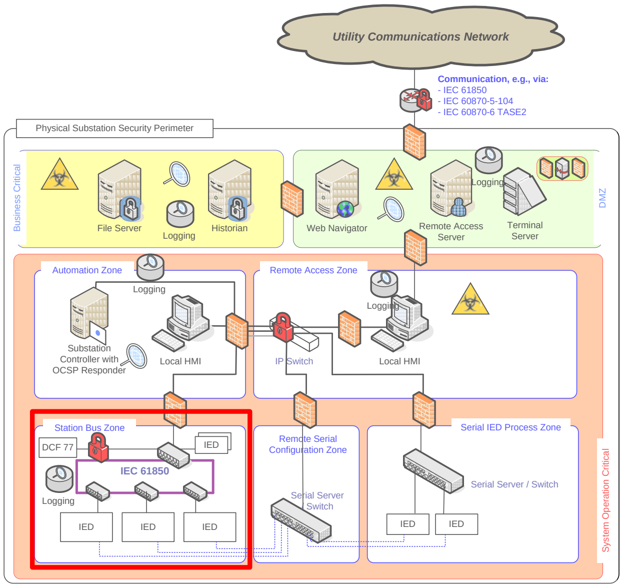


Figure 2. Substation – Functional Split into Zones

As depicted in Figure 2, the substation bus can be realized as ring, connecting the protection relays, acting in real-time. There is a connection to other zones within the substation, separated from the real-time part using Firewalls. Examples are the automation zone or the remote access zone. Another example is the zone storing the historian information also interacting with a backend SCADA system. The historian is a device for archiving measurements, events, and alarms of the substation. Figure 2 already shows security elements deployed within a substation, like Firewalls, virus checking tools, or access control means to components or data.

### B. *DER Incorporation*

Decentralized Energy Resources (DER) may be connected to the Smart Grid at two different connection points. Depending on the amount of energy provided, they may be connected to the low voltage network or to the medium voltage network (distribution network). The first one is rather typical for DER in residential areas, like a solar panel, while the connection to the medium voltage network is done for larger deployments like wind power farms or solar parks. Necessary for both is the connectivity to a communication infrastructure to allow a control center to act on provided information about current energy generation, but also to provide scheduling information to the DER, e.g., depending on the weather forecast, to better balance the feed in of energy into the electrical network. Communication with the DER may be done using different communication technologies, like Power Line Communication (PLC) or wireless communication via the UMTS network. For the distribution network operator (DNO), it is essential to know, which DERs are associated to his operational control. This can be supported by the used security credentials using additional information depending, e.g., on the geographic location of a DER or on the association with a dedicated DNO.

## III. Secure Communication in Smart Grid

IEC 61850 [3], [4] is a standard for communication in the domain of energy automation. It is envisaged to be the successor of the currently used standards IEC 60870-5-104 and DNP3 especially used in the North American region. IEC 61850 enables interoperability between devices used in energy automation. For example, two IEC 61850 enabled devices of different manufacturers can exchange a set of clearly defined data, and the devices can interpret and use these data to achieve the functionality required by the application due to a standardized data model. In particular, IEC 61850 enables continuous communication from a control station to decentralized energy generators or to IEDs (like protection relays) in a substation.

IT security is increasingly important in energy automation as on part of the Smart Grid. Here, the IEC 62351 framework [5] with currently 11 parts kicks in, defining security services for IEC 61850 based communication covering different deployment scenarios using serial communication, IP-based communication, and also Ethernet communication. The latter one is used locally within a substation to cope with the high real-time requirements. While it may be not always necessary to encrypt the communication to protect confidentiality, there is a high demand to protect the communication against manipulation and to allow for source authentication. IEC 62351 relies on existing security technologies as much as possible and profiles it for the application environment . One example is the application of Transport Layer Security (TLS, RFC 5246 [6]) to protect TCP-based communication. Here, IEC 62351 basically reduces the manifold options of TLS to ease interoperability. Another example is the adoption of Group Domain of Interpretation (GDOI, RFC 6407 [7]) as group-based key management to distribute key material for the protection of status information and event signaling between IEDs in a substation or across substations using Wide Area Networks (WANs).

A specific characteristic throughout IEC 62351 is the consequent application of X.509 certificates and corresponding private keys for mutual authentication on network layer and application layer. This requires an efficient handling of X.509 key material and the availability of this information right from the installation. There is a strong need to provide these credentials without increasing the installation effort. For instance, devices may generate their own key pair, but certification needs to bind this key pair to the operational. This is a challenge from the pure technical perspective as a high number of devices need to be equipped with the key material. But also from the network operator process point of view this is challenging, as the key material has a lifecycle and needs to be updated once in a while. These aspects will be addressed in the following sections.

## IV. Public Key Infrastructure - PKI

A PKI typically contains a variety of services requiring interfaces in the devices utilizing the PKI and also an accompanying process. In general, a Public Key Infrastructure provides a secure, reliable, and scalable environment for the complete lifecycle of key material, i.e., generating, distributing and querying public keys for secrecy, correctness, and sender verification. Moreover, it binds the “owner” to the public key using a digital certificate and thus enables identification of users and components utilizing certificates. Furthermore, it maintains and distributes status information for the lifetime of that binding, i.e., from the generation till the revocation. The general functionality and formats are described in RFC 5280 [8].

The following list provides a short overview about the different components, which are depicted in Figure 3:

- **Registration authority (RA)** authenticates the user or IED or the data submitted by the user or IED, performs an authorization check and initiates the certificate generation at the CA. For machine-to machine communication the RA can be used to mediate between the device applying for a certificate and the CA.
- **Certification authority (CA)** is a trusted entity that certifies public-keys by issuing certificates.
- **Key/certificate archive** is a repository in which the CA stores certificates and/or generated key pairs.
- **Key generation** is a function of the PKI responsible for the generation of key material (public and private keys), which are certified through the CA
- **Public Directory** is a (usually publicly readable) database to which the CA stores all issued certificates
- **Revocation Lists** are also a publicly readable database to which the CA stores all revoked certificates

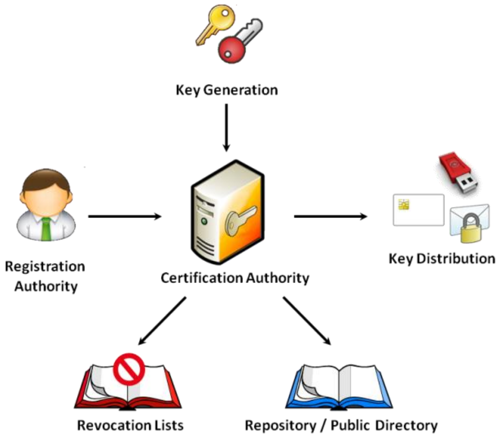


Figure 3. PKI Components

In the context of smart grid, a PKI may be operated by a utility company working as internal PKI, or it may be a public PKI, also depending on the target use case and the need for interoperation between different parties. Moreover, the functionality provided by the PKI needs to be streamlined to the target environment to avoid unnecessary effort. In any case, the devices utilizing key material issued by the CA need to provide the technical interfaces to accomplish this task. This is described in more detail in IEC 62351-9 targeting the key management explicitly.

Section VI describes an enhancement of the typical used PKI setup by introducing an intermediary, which provides all operational environment specific information. This avoids the pre-configuration of IEDs with this information.

## V. EXISTING CERTIFICATE ENROLLMENT METHODS

This section describes common methods for certificate enrollment taking device capabilities into account. Capabilities in this context relate to local and remote key generation. Typically local generation of key material is desired to avoid the handling of private keys outside the devices. Note that depending on the key usage, there may be requirements to also have the private key available in a trust center to ensure that encrypted information can be accessed even the device hosting the private key was either damaged or has been compromised.

### A. Manual Enrollment

Manual enrollment relates to the manual connection of a device to an engineering tool to provide the key material during a local configuration session, prior to the connection in the target network. This approach requires a significant initial configuration effort and is especially cumbersome in case of device replacements. It may be realized using an offline engineering network to bootstrap the security credentials for connected devices. Here, the devices or components do not possess a cryptographic credential up front, as the separate network is assumed to be physically secure. In the simplest form, it may be a direct connection of an engineering laptop to the component to be administered using purely local point-to-point communication.

### B. Automated Enrollment

Automated enrollment refers to the initial configuration of devices including the key material. This is shown in Figure 4. Field devices are connected to the network and contact the PKI server to obtain certified key material.

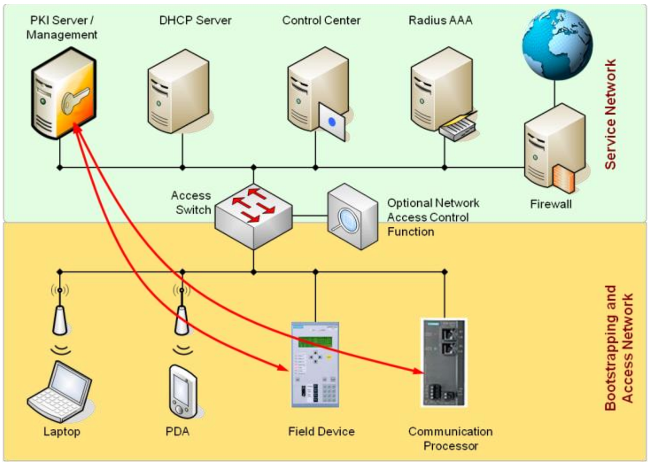


Figure 4: Automated distribution using management protocols

Here, the field devices generate their public/private key pairs locally and send a Certificate Signing Request (CSR) for the public key to the RA/CA (part of the PKI server). Part of the CSR may be a serial number of the device, against which the PKI server can check a configured list of devices allowed to be enrolled. This authorization may also be realized by other means like one-time passwords. According to RFC 2986 [9] the CSR is defined n ASN.1 as shown in Figure 5 below.

```
CertificationRequest ::= SEQUENCE {
   certificationRequestInfo
     CertificationRequestInfo,
   signatureAlgorithm AlgorithmIdentifier{{
     SignatureAlgorithms }},
   Signature BIT STRING
   }

CertificationRequestInfo:
   CertificationRequestInfo ::= SEQUENCE {
   version       INTEGER { v1(0) } (v1,...),
   subject       Name,
   subjectPKInfo SubjectPublicKeyInfo{{
    PKInfoAlgorithms }},
   Attributes [0] Attributes{{
    CRIAttributes }}
   }

SubjectPublicKeyInfo { ALGORITHM : IOSet}
    ::= SEQUENCE {
    Algorithm AlgorithmIdentifier {{IOSet}},
    subjectPublicKey BIT STRING
   }
```

Figure 5: Certification Request structure [9]

Several protocols are known for transmitting a CSR to a CA. Examples are:

- SCEP – Simple Certificate Enrollment Protocol [10]
- CMP – Certificate Management Protocol [11]
- CMC – Certificate Management over CMC [12]
- EST – Enrollment over Secure Transport [13]
- XML Key Management Specification [14]

These different protocols describe the communication of a CSR from a device to the CA, were the device ideally generates the key pair for itself. Additionally to identification information like the serial number, further information can be connected with the CSR, like a password (to be used to authorize a potential future revocation) or key usage restrictions. The CSR has to be protected to prevent illegitimate issuing of certificates. The CSR itself may be protected using the public key of the RA/CA as in case of SCEP. In case of CMP, the CSR is protected using an initial authentication key, and in EST, the CSR is transmitted over a secured communication link. Here TLS is applied, providing the opportunity to authenticate both peers during the connection establishment. Also, there may be an intermediate RA located between the device sending the CSR and the CA, which already performs the verification of the CSR to reduce the load on the CA.

When deployed in the operational environment, IEDs are typically not pre-configured. Hence, an intermediate component is used to enhance the CSR with additional information about the deployment environment before it is forwarded to the RA/CA. This information is not available at or provided by the sender of the CSR itself. The following section describes such an enhancement of the CSR communication on the way from the devices to certification server. This enhancement is proposed to provide additional information about the environment in which the device is deployed. Such information can either be contained in the certificate to be issued or associated with the device certificate by other means, like a central configuration database. This approach helps identifying, e.g., a physical movement of components or devices to other locations. Hence, key material valid in one location may not be misused in a different location. Moreover, the approach also enhances the options for asset management, by providing fine-grained information already during the authentication processes, employing the enhanced certificate.

## VI. Enhancing Certificate Enrollment with Device Purpose Binding

This section outlines the introduction of an additional network component to extend a CSR with additional information. Such additional information is encoded as additional attribute added to the original CSR as sent by the device. This attribute indicates the context or other deployment specific information, to be added to either the certificate or the configuration database.

This is achieved by adding a Certificate Attribute Intermediary (CAI) along the CSR communication path. The CAI adds at least one attribute to the original CSR (without otherwise manipulating the original CSR). The additional attribute acknowledges additional information about the operating environment. This additional information may be the membership of the CSR sender (device) to a dedicated zone or group or to a dedicated location either on a geographical base or on an organizational base. Moreover, the CAI may already check the CSR (like an RA) and signal this also in the attribute. The CAI may add information about intended usage restrictions of the certificate, depending on the device type and the security policy. This information can be part of the engineering information, which must then be available at the CAI. The CAI may also request that the certificate is issued using a dedicated signature algorithm.

The attribute and the CSR build the Extended Certificate Request (ECR). The ECR is protected by a cryptographic checksum, binding the attributes to the original CSR. Ideally, this is a digital signature of the CAI. This could be realized as PKCS#7 structure [15] or as XML structure, but may also be a symmetric checksum, involving a shared secret between the CAI and the CA. The ECR is then forwarded to the RA/CA, which verifies both the CSR and the additional attribute. If the RA and CA are separate entities, the CAI may be co-located with a local RA. After successful verification, the additional information from the attribute is included in the X.509 certificate within a certificate extension.

Depending on the applied enrollment protocol the ECR may be transmitted via a TLS protected communication path using, e.g., HTTP POST, HTTP GET or as REST or SOAP message).

Figure 6 depicts the on path enhancement of a CSR with attributes *aa1, ..., aa3*. Also shown are potential functions to be performed by the CAI (e.g., CSR checking) and the enhanced functions on the RA/CA side.

In a substation automation environment, the CAI can be part the substation controller or the remote access server as the central ingress and egress point of the substation. This is depicted in Figure 7. The different steps describe the single steps for the ECR processing. Note that the prerequisite is the availability of the central RA/CA root certificate in the IED.

The following steps are performed for the initial enrollment of an energy automation device IED:

1. Generation of key material (public/private key), generation of the CSR within the IED
2. Send local generated CSR to Remote Access Server
3. Verification of CSR through Remote Access Server. Remote Access Server acts as CAI. Generation of attributes and ECR. Send ECR to central RA/CA server of the distribution network operator.
4. Verification of ECR signature through central RA/CA, verification of attributes (installation information, etc.); optional verification of original CSR
5. In case of successful verification device specific certificate will be generated and send to the remote access server of the substation.
6. Forwarding of certificate to IED
7. Local automated installation of certificate upon receiving and successful signature verification against local RA/CA root certificate.

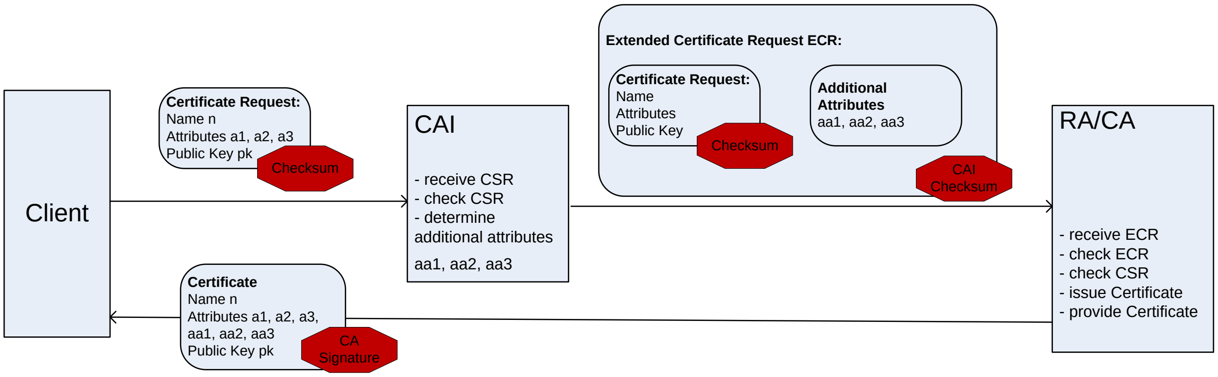


Figure 6: Realization option for on path CSR enhancement with attributes characterizing the deployment environment

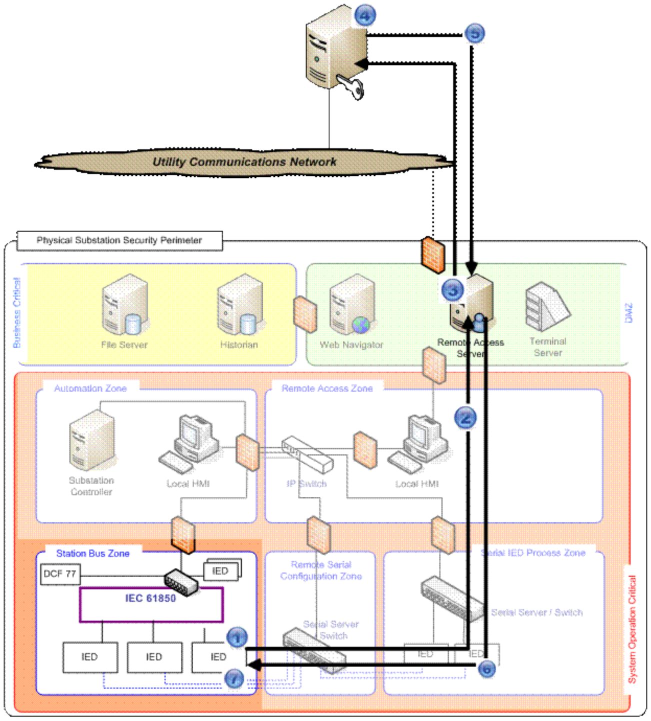


Figure 7: Enhancement of the CSR path in a substation

Note that this paper describes the concept of the enhancement. Implementations are not finished yet.

## VII. Conclusions and Outlook

This paper described security enhancements for energy automation systems involved in substation communication and smart grid. The cryptographic protection of control communication requires that cryptographic keys and certificates are provisioned on energy automation devices. Manual configuration would not scale to the huge number of devices, and be prone to configuration errors. Therefore, automatic configuration of automation devices is required not only during the operation, but especially for the initial device enrollment. To ensure the correct configuration of cryptographic device credentials, information is required at which location a specific device has been installed. An additional network element has been described that trustfully enhances a certificate signing request issued by an automation device with information on the network segment in which the device has been installed. This allows the CA to issue a device certificate that is bound to the operational zone of the device ("location"). Moreover, additional information for the CSR processing can also be provided. A relying device towards which the considered device authenticates using this zone-bound certificate, can verify whether the device belongs to the own zone. This ensures that an automatically provisioned device is operable using the established configuration only within the corresponding zone. When the device is relocated or put out of service, its device certificate cannot be misused, e.g., in other zones.

Standardization is currently ongoing in the context of ISO/IEC62351-9, which defines interoperable means for automatic device credential management for energy automation equipment. The new approach described in this paper enhances the current credential management approach and will be proposed for to be considered in future energy automation security standards.

## References

[1] S. Fries and R. Falk, "Efficient Multicast Authentication in Energy Environments", Proc. IARIA Energy 2013, March 2013, ISBN 978-1-61208-259-2, pp. 65-71, http://www.thinkmind.org/download.php?articleid=energy_2013_3_30_40056 [retrieved Dec. 2013]

[2] M. Felser, "Real-time Ethernet – industry prospective," Proc. IEEE, vol. 93, no.6, June 2005, pp. 1118-1128, http://www.felser.ch/download/FE-TR-0507.pdf [retrieved: Dec. 2013]

[3] IEC 61850-5 – "Communication requirements for functions and device models", July 2003, http://www.iec.ch/smartgrid/standards/.

[4] "Efficient Energy Automation with the IEC 61850 Standard Application Examples", Siemens AG, December 2010, http://www.energy.siemens.com/mx/pool/hq/energy-topics/standards/iec-61850/Application_examples_en.pdf [retrieved: Dec. 2013].

[5] IEC 62351-x Power systems management and associated information exchange – Data and communication security, http://www.iec.ch/smartgrid/standards/.

[6] T. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, Aug. 2008, http://tools.ietf.org/html/rfc5246 [retrieved: Jan. 2014].

[7] B. Weiss, S. Rowles, and T. Hardjono, "The Group Domain of Interpretation", RFC 6407, Oct. 2011, http://tools.ietf.org/html/rfc6407 [retrieved: Jan. 2014].

[8] D. Cooper et al, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile", RFC 5280, May 2008, http://tools.ietf.org/html/rfc5280 [retrieved: Jan. 2014].

[9] M. Nystrom and B. Kaliski, "PKCS #10: Certification Request Syntax Specification", RFC 2986, Nov. 2000, http://tools.ietf.org/html/rfc2986 [retrieved: Jan. 2014].

[10] M. Pritikin, A. Nourse, and J. Vilhuber, "Simple Certificate Enrolment Protocol", Internet Draft, Sep. 2011, http://tools.ietf.org/html/draft-nourse-scep-23 [retrieved: Jan. 2014].

[11] J.Schaad and M.Myers, "Certificate Management over CMS", RFC 5272, June 2008, http://tools.ietf.org/search/rfc5272 [retrieved: Jan. 2014].

[12] C. Adams, S. Farrell, T. Kause, and T. Mononen, "Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP), RFC 4210, Sep. 2005, http://tools.ietf.org/html/rfc4210 [retrieved: Jan. 2014].

[13] M. Pritikin, P. Yee, and D. Harkins, "Enrollment over Secure Transport", RFC 7030, Oct. 2013, http://tools.ietf.org/html/rfc7030 [retrieved: Jan. 2014].

[14] XML Key Management Specification http://www.w3.org/TR/xkms2/

[15] B. Kaliski, "PKCS#7 Cryptographic Message Syntax Version 1.5, RFC2315, March 1998, http://tools.ietf.org/html/rfc2315 [retrieved: Dec. 2013].

# An Efficient Search Scheme Based on Perfect Difference Graph for P2P Networks

Chen-Wei Wang and Yaw-Chung Chen
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
{iverson2220@yahoo.com.tw, ycchen@cs.nctu.edu.tw}

***Abstract*** **- We propose an efficient search scheme for multi-layer unstructured P2P systems, and show that it is not only reliable, but also scalable. To the best of our knowledge, there are few researches focusing on the reliable and scalable search mechanism for unstructured P2P systems. The broadcasting performance of the P2P system is enhanced through the use of a Multi-hop Index Replication with Perfect Difference Graph (PDG) forwarding algorithm, which makes certain that each super-peer receives just one copy of the broadcast message. Furthermore, by using the Multi-hop Index, a super-peer has extra information to know whether the queried file is available or not. The experimental results show that our proposed scheme improves existing unstructured P2P systems in terms of a higher query success ratio, fewer query flooding messages and shorter average delays. In other words, our proposed scheme achieves high scalability, low communication overhead and improved performance of query responses.**

***Keywords: Perfect Difference Graph; P2P systems; Multihop-index***

## I. INTRODUCTION

Structured P2P overlay networks can provide efficient and accurate query service but it needs a large overhead to maintain the distributed hash table (DHT) and suffers peer churn. Measurement studies of deployed P2P overlays [19][20] show a high rate of churn. Unstructured P2P overlays use flooding or random walks to look up data items. It is resilient to churn, but its queries may generate a large volume of traffic and cause systems un-scalable. Although queries by using random walk reduce the traffic volume, huge traffic overhead still occurs when the requested resource does not exist. Unfortunately, this situation happens often. In [7] authors observed that roughly a half of the queries cannot be matched. A solution to reduce the query traffic is setting the time-to-live (TTL) of query packets to a small value. However, this approach will search only a small portion of the peers in the system and queries are very likely failed even if the requested resource does exist. Measurement studies on actual unstructured P2P networks observed that the ratio of successful query is typically around 10% [21].

In this paper, we propose Multi-hop Index Replication that can improve search performance for rare objects. Index replication features not only much lower overhead compared with data replication, but also more effective for improving the scalability of unstructured networks [6][11][20]. Our Multi-hop Index Replication with PDG forwarding algorithm eliminates the impacts of redundant query flooding messages and reduces the traffic in searching unavailable files. We use a so-called bootstrap peer (BSP) to maintain a super-peer table, a peer joining the P2P network and wishing to become a super-peer must first send a request to the BSP. After examining the requesting peer's bandwidth conditions, the BSP may select the peer as a super-peer, and send it the corresponding connection information or register the peer as a redundant super-peer.

When the overlay topology is established, a pure PDG [8][9] forwarding algorithm can be used to transport the query messages from the originating super-peer to other super-peers in such a way that each super-peer receives just one copy of the message. In addition, each super-peer has to maintain an AVL tree-based index that is constructed with a randomly generated key. The average-case complexity of search, insert, and delete operations is O(log $n$), where $n$ is the number of shared files in the overlay network.

The rest of the paper is organized as follows. In Section II we present the background related to both structured P2P and unstructured P2P systems. In Section III, we discuss the proposed hybrid P2P system with analytic models. In Section IV, we evaluate the effectiveness of our methods and discuss the performance of the system. We conclude the work and address the future issues in Section V.

## II. BACKGROUND

Peer-to-peer (P2P) architecture is an alternative to the traditional client/server architecture. P2P makes it possible for users to organize themselves into ad hoc groups that can efficiently and securely handle requests, share resources, collaborate, and communicate. As P2P systems evolve, we can anticipate the emergence of a wide variety of online communities [18].

### *A. P2P File Sharing Applications*

Many P2P systems have been proposed for different applications [1]-[6], [10]-[14]. In this paper, we focus on P2P overlays for efficient data (file) sharing among peers. The Content Addressable Network (CAN) [15] was proposed for file sharing and the entire space is partitioned to distinct zones such that each peer is in charge of one zone. Every peer maintains a routing table which holds the IP address of its neighbors in the coordinate space. The data is stored into and retrieved from the peer that owns the zone. CAN takes advantage of the ordering of the Cartesian coordinate space in the routing protocol.

Chord [10] organizes the node keys into a so–called chord ring, where each peer is assigned an ID. Peers are inserted into the ring according to the order of their IDs. Each peer has a successor and a predecessor. To accelerate the search, each peer maintains a finger table, in which each finger points to a peer with a certain distance from the current peer. Compared to CAN, Chord is simpler as the key is hashed into a one dimensional space.

Gnutella [12] is a decentralized unstructured peer-to-peer overlay, in which peers join the system based on loose rules. To look up a data item, a peer sends a flooding search request to all neighbors within certain radius. In Gnutella, flooding method consumes a lot of network bandwidth and hence the system is not scalable. Also, it is usually hard to find a rare data item because it is unlikely to flood the search request to all peers. The study in [18] improved the efficiency while looking up a rare data item.

BitTorrent[13] is a centralized unstructured P2P network. It uses a central server called tracker to keep track of peers in which files are stored. The tracker records the network location of each client either uploading or downloading the file associated with a torrent. Each file has a corresponding torrent file stored in the tracker. Upon receiving a download request, the tracker sends back a random list of peers which with the same file. BitTorrent suffers single point of failure problem at the central server.

YAPPERS [15] combines both structured and unstructured P2P overlays to provide a scalable search service over an arbitrary topology. It is designed for efficient partial search which only returns partial values of data. For a complete search, YAPPERS still needs to flood the query to all peers which are in the same color as the data. Compared to YAPPERS, our proposed multi-layer unstructured P2P system can further improve the accuracy of the lookups in a more efficient way.

In [16], the structured overlay was used to support unsuccessful flooding data search, however, its structured overlay is responsible for connecting all the unstructured overlays and transferring query requests between them.

### B. The P2P Lookup Problem

In addition to single point of failure and poor scalability, a drawback of centralized approach is the vulnerability to malicious attacks and legality issue. These shortcomings led to the adoption of decentralized solutions, which make it difficult to ensure high performance and availability, so a high degree of redundancy is required.

## III. PROPOSED APPROACHES

Neither structured P2P nor unstructured P2P networks alone can fulfill the requirements of efficiency, scalability, and reliability of services. The motivation of this study is to combine both types of P2P networks to provide a hybrid approach which can offer better efficiency and scalability.

### A. Super-Peer Overlays and Forwarding Protocols

The super-peers overlay topology can be constructed as a graph, in which vertices represent individual super-peers while undirected edges stand for connections between super-peers. As shown in Table I, each peer in perfect difference graph (PDG) [8] has a degree $O\ (\sqrt{n})$, thus the topology is more flexible than $O(n)$ in the complete graph. Furthermore, the search range of a PDG-based topology is similar to that in the complete graph-based topology. Table I shows that other graph topologies have both a lower peer connection degree and a larger diameter (i.e. no. of hops along the path) than the PDG approach. Thus, the PDG-based overlay topology is a better choice for the hybrid P2P system presented in this paper.

#### 1) Perfect difference graph:

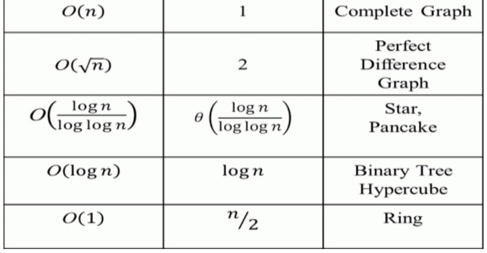
According to the definition of perfect difference sets (PDSs), PDG provides the mathematical knowledge for achieving this optimum number of peers to construct the framework of perfect difference networks or PDNs as defined bellow:

**Definition 1:** Perfect Difference Network (PDN) — there are $n = \delta^2 + \delta + 1$ nodes, numbered 0 to $n$-1. Node $i$ is connected to node $i \pm 1$ and $i \pm s_j (mod\ n)$, for $2 \leq j \leq \delta$, where $s_j$ is an element of the PDS $\{s_0, s_1, \ldots, s_\delta\}$ of order $\delta$.

Table II illustrates the number of peers, the number of elements and the order in the first ten PDSs. Fig. 1 shows a PDG overlay based on the PDS {1, 3}. Since there are two elements in the PDS, the graph has seven ($2^2 + 2 + 1$ = 7) peers. For example, peer 0 has edges connecting to peers (0±1) mod 7 and (0±3) mod 7, which are peer 1, 3, 4 and 6.

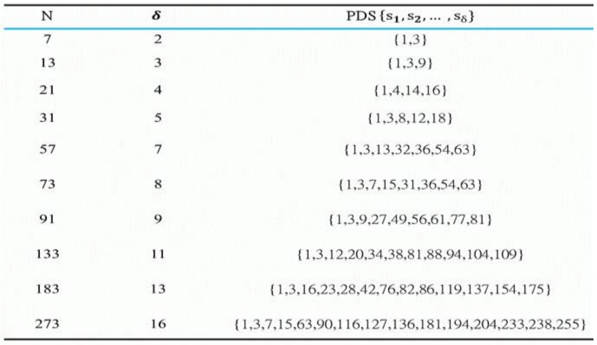
For example, in Fig. 1, we present a brief description of the forward edges of peer 0, which are the edges connecting peer 0 to peer 1 and 3, respectively, and the backward edges are the edges connecting peer 0 to peer 4 and 6, respectively.

TABLE I. COMPARISON OF VERTEX DEGREE AND GRAPH DIAMETER.



| Order of vertex degree | Graph diameter | Example network |
|---|---|---|
| $O(n)$ | 1 | Complete Graph |
| $O(\sqrt{n})$ | 2 | Perfect Difference Graph |
| $O\left(\frac{\log n}{\log\log n}\right)$ | $\theta\left(\frac{\log n}{\log\log n}\right)$ | Star, Pancake |
| $O(\log n)$ | $\log n$ | Binary Tree<br>Hypercube |
| $O(1)$ | $^{n}/_{2}$ | Ring |

TABLE II. RELATION BETWEEN NUMBER OF SUPER-PEERS N, ORDER Δ AND PDS



| N | $\delta$ | PDS $\{s_1, s_2, \dots, s_\delta\}$ |
|---|---|---|
| 7 | 2 | {1,3} |
| 13 | 3 | {1,3,9} |
| 21 | 4 | {1,4,14,16} |
| 31 | 5 | {1,3,8,12,18} |
| 57 | 7 | {1,3,13,32,36,54,63} |
| 73 | 8 | {1,3,7,15,31,36,54,63} |
| 91 | 9 | {1,3,9,27,49,56,61,77,81} |
| 133 | 11 | {1,3,12,20,34,38,81,88,94,104,109} |
| 183 | 13 | {1,3,16,23,28,42,76,82,86,119,137,154,175} |
| 273 | 16 | {1,3,7,15,63,90,116,127,136,181,194,204,233,238,255} |

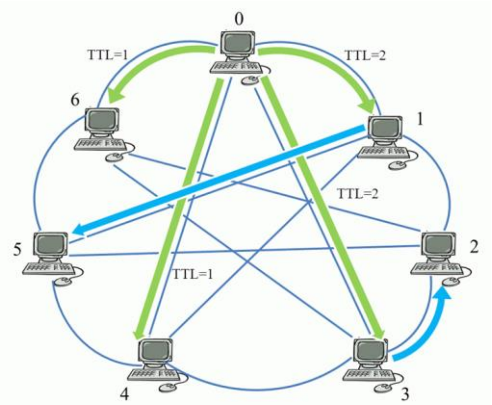


Figure 1. An example of PDG-based forwarding algorithm.

### 2) Broadcasting over a Super-peer Overlay

We deploy a PDG-based forwarding algorithm [9] in which the query requests are delivered to all super-peers in the overlay via the forward and backward edges of the perfect difference network. Each super-peer will send the search requests by using the PDG forwarding algorithm so that each super-peer receives only one copy of the search requests. The PDG forward query message in two steps:

- **Step 1**: Super-peer *i* sends a request message with TTL=2 to all of its forward partners and sends a request message with TTL=1 to all of its backward partners.
- **Step 2**: If an intermediate super-peer receives the request message, it duplicates the message to all of its backward partners other than the partner from which it received the original query message.

Fig. 1 illustrates the PDG-based forwarding algorithm for a super-peer overlay forming a PDG with an order of $\delta = 2$. Here super-peer 0 wants to send a query to all other super-peers. According to the above two steps, super-peer 0 sends a query request with TTL=1 and 2 by its forward and backward edges to partners {1, 3} and {4, 6}, respectively. In the case of TTL=2, the TTL value is reduced to 1, and partners {1, 3} forward a copy of the query request to all their backward peers other than the edge from which they received the query message. In the case of TTL=1, since the TTL value is reduced to zero, partners 4 and 6 take no further action.

### 3) Multi-hop Index Replication

Since each super-peer maintains the index replication in AVL tree, so that lookup, insertion, and deletion all take O(log *n*) time in both the average and worst cases, where *n* is the number of shared files in the overlay.

To ease the discussion, we explore two-hop index replication strategy in which each ordinary peer sends the name of shared files to all of its one-hop super-peer, which maintains the index replication in AVL tree. The index is constructed with a randomly generated key that is the name of shared files published by the ordinary peers through the SHA1-like algorithm. These hashed keys are inserted to the AVL tree-based index. Each super-peer broadcasts the available resource names by using PDG algorithm. In order to further reducing the broadcast messages, each super-peer uses 1-bit to record the status of the shared files. If the bit were set, the shared file would come from ordinary peers it controlled. Otherwise the file would be in other super-peers. By this way, only when the bit is set, does a super-peer sends the query messages to its ordinary peers. Otherwise, the super-peer will forward the lookup messages to other super-peers directly by PDG-based algorithm. Each super-peer can search the AVL tree to decide whether to broadcast messages to its ordinary peers or not. If we can't find any records in the AVL tree-based index, it means that there is no resource published by peers.

## B. System Construction and Architecture

The bootstrap peer (BSP) uses a super-peer table to maintain the super-peer overlay structure. For convenience, we discuss only one bootstrap peer attached to the overlay network.

### 1) System Construction

In our example system, the ordinary peer can connect to two super-peers. Thus when one of the super-peers leaves or crashes, the other can still hold the records. When a new peer enters the overlay as an ordinary peer, it sends a request to the bootstrap peer. Upon receiving the joining request, BSP acknowledges the peer with a list containing the IP addresses of randomly selected super-peers. The peer then chooses a less loaded super-peer to establish a connection. Once the new peer connects to the super-peer, it becomes one of ordinary peers of that super-peer and super-peer sends it a peer list which contains the part of the ordinary peers in the same overlay. When the ordinary peer wants to leave the system, it simply sends a message to inform its super-peer, which then updates the corresponding AVL tree-based index to show that the shared files in the leaving peer no longer exist in the tree. Fig. 2 shows the overlay configuration in which a new peer is joining the P2P overlay. By multi-layer and multi-hop architecture, our system can serve as much powerful super-peer (MSP) with large storage space, computational capability and higher bandwidth to manage the whole PDN cluster, as shown in Fig. 3.

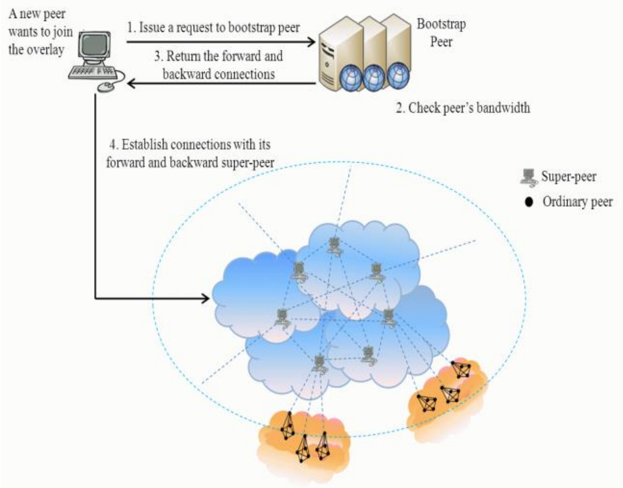

A new peer wants to join the overlay
1. Issue a request to bootstrap peer
3. Return the forward and backward connections
Bootstrap Peer
2. Check peer's bandwidth
4. Establish connections with its forward and backward super-peer
Super-peer
Ordinary peer

Figure 2. Configuration of a new peer joining the super-peer.

*2) New Super-peer Joining*

Any peer with high bandwidth entering the P2P overlay network sends a joining request with its bandwidth and IP information to the BSP. After checking the bandwidth quality, the BSP may accept the peer as a super-peer. When the number of super-peers is larger than the value $(\delta^2 + \delta + 1)$, a newly joining peer that fulfills the bandwidth requirement will be marked by BSP as a redundant super-peer. When the total number of super-peers increases to a threshold, $1/2\,[(\delta^2 + \delta) + (l^2 + l)]$, the current PDS will be extended to the successor PDS order and the super-peer overlay will be extended accordingly.

In the initial set-up phase, the BSP utilizes a min-order PDS with order 2 to construct a basic super-peer overlay network for a maximum of 7 super-peers. Assume that 9 new peers fulfill the bandwidth requirements as a super-peer, since the number of new peers exceeds the number of available spaces in the overlay, the former 7 peers are assigned as super-peers, while the remaining peers are appointed as redundant peers. Later, when a newly coming peer wants to become a super-peer, it will result in the total number of super-peers, including active, new incoming, and redundant super-peer, exceeding the threshold 9(= (6+12)/2). The BSP then extends the super-peer overlay topology using a PDS with an order of 3, thus allowing for super-peers up to 13, including 10 active super peers in the newly extended configuration.

*3) Super-peer Leaveing*

When a super-peer is leaving the system, it sends a leave message to both the BSP and all of its ordinary children peers. The BSP selects one of the redundant super-peers to take place the leaving one. Then BSP sets the active state of the redundant peer and informs other active peers to update their partner records correspondingly.

Having received a leave message from a super-peer that tries to disconnect from the P2P overlay, each ordinary peers re-enter to the overlay by choosing one of the super-peers with the shortest response time in its super-peer list. When the number of super-peers is below the threshold $(\delta^2 + \delta + 1)$, there will be no enough super-peers to take over the leaving peers in the overlay network, some of the super-peers lose their forward or backward partners. As a result, some of the super-peers may fail to receive the messages delivered by other super-peers in the overlay. To overcome this effect, when the number of super-peers decreases to the threshold, $1/2\,[(\delta^2 + \delta) + (l^2 + l)]$, the order of the current PDS will be reduced to the predecessor of the PDS, and the super-peer overlay topology will be reduced consequently. Then BSP computes and updates new forward and backward partners based on the new order $\delta$ in its super-peer table and sets the status of those redundant super-peers to 1. Finally, the BSP notifies active super-peers of the forward and backward partners and redundant super-peers.

Assuming 10 active super-peers in a super-peer overlay use a PDS with an order of 3. If one active super-peer sends a leaving message, the BSP reduces the topology because the number of super-peers equals the threshold 9(= (6+12)/2). The BSP appoints a min-order PDS (an order of 2) to reduce the current super-peer overlay, thus allowing up to 7 super-peers. It assigns new peer ID to the remaining super-peers. The super-peers with peer ID less than 7 are appointed as active super-peers in the reduced topology. The rest are appointed as redundant super-peers. As a result, 7 active super-peers and 2 redundant super-peers exist in the system.

## C. Quantitive Evaluation

We evaluated the existing mechanisms in terms of the query success ratio, number of flooded lookup query messages and average delay. All the results are the average of 10 simulation runs. Let $p_a$ be the probability of the resource recorded in the AVL tree and $\tau$ the connection degree. For a two-layer unstructured P2P system, the super-peer layer is in mesh-based structure, the location of a data item is arbitrary, and it uses flooding to do a best-effort search. With pure-PDG structure, whether the resources exist or not, it always sends the query messages,

$$p_a \times \tau + (1 - p_a) \times (1 + \tau + \tau^2) \qquad (1)$$

For super-peer layer with AVL-PDG structure, we want to further reduce the number of lookup messages between super-peers. The range of the flooding is

$$\begin{cases} 1 & \text{, if the resource can't be matched} \\ \mathrm{p_a} \times \tau & \text{, if the resource is in the same local area} \\ (1 - p_a)(1 + \tau + \tau^2) & \text{, if the resource is in other overlay} \end{cases} \qquad (2)$$

## IV. Simulation and Numerical Results

We present the simulation results to evaluate the performance of the random mesh-based, hierarchical unstructured P2P system and our proposed scheme. From various aspects of performance, we show that our proposed approach is practical and works well.

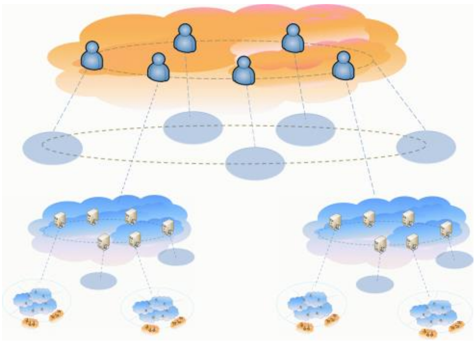


Figure 3. Multi-layer architecture for storage system.

### A. *Simulation Environment and Simulation Setup*

We perform simulation with NS-2 (version 2.27) simulation tool [14] with GnutellaSim to evaluate our proposed method. We adopt Gnutella as our basic architecture in the simulation and add our proposed approach to the basic scheme. We validate the performance and improvement through the simulation result.

#### 1) *Simulation Environment*

The topology used for simulation is presented in Fig. 2. It consists of a bootstrap peer (BSP), super-peers and ordinary peers. A new peer entering the overlay will send a request to the BSP, which then acknowledges the peer with a list containing the addresses of randomly selected super-peers. Then the joining peer starts to query the resources it needs. The general parameters are presented in Table III.

#### 2) *Simulation Setup*

Each network topology is composed of 1,000 nodes, and each node is assigned as either a super-peer or an ordinary peer randomly. The ratio between the number of super-peers and the total number of peers is set to 10%. We set super-peer's download bandwidth to at least 2 Mbps and upload bandwidth at least 1.5 Mbps. The ordinary peer's download bandwidth is no more than 1.5 Mbps and connection degree is 2. We set the peers with heterogeneous link capacities such that 10% of the peers have the link capacity greater than 2Mbps, 30% of them have the link capacity less than 1.5Mbps, and 60% of them have the link capacity between these two values. Each connection's link delay is randomly set between 1 and 10 ms. We run the simulation 10 times to get its average and the duration is 1000 seconds each time.

TABLE III. EXPERIMENTAL ENVIRONMENT.

| **Parameter** | **Value** |
|---|---|
| Peer Number | 5,000 |
| Simulation Time | 1,000(sec) |
| Number of files | 100,000 |
| Query frequency | 5 per sec |
| Super-peer Download Bandwidth | > 2 Mbps |
| Super-peer Upload Bandwidth | > 1.5 Mbps |
| Peer Max degree | 2 |
| Connection link delay | 1~10(ms) |

### B. *Numerical Results*

For our simulations, we modified an implementation of Gnutella [12].

#### 1) *Aerage Traffic*

Fig. 4 shows that the AVL-list with PDG overlay has the lightest broadcast overhead. In Fig. 5, the results clearly demonstrate the success rate of the AVL-list with PDG overlay significantly higher than mesh-based overlay and pure PDG overlay. The AVL-list improves about 40% and 50% of query success ratio compared to the pure-PDG and mesh-based overlay, respectively as shown in Table IV.

#### 2) *Performance of the Network Traffic*

Fig. 6 shows the comparison of network traffic between mesh-based overlay and PDG-based overlay. In the former, queries for unavailable files can generate an unbounded traffic load. While the traffic load with AVL-list is bounded.

#### 3) *Performance of the Response Time*

Fig. 7 shows that PDG forwarding can help resolve the queries in the super-peer overlay, which has smaller diameter than the entire P2P network. The average response time is kept between 1.5 and 2 hops with 5000 peers in the system. The simulation result shows that the average response time decreases from 4.3 hops to 1.7 hops. The overall user perceived response time can be reduced by 60%.

## V. CONCLUSION AND FUTURE WORKS

We propose a novel and efficient search scheme for multi-layer P2P systems using a Multi-hop Index Replication with PDG forwarding algorithm. We show that our scheme is not only reliable but also scalable. Our work shows that unstructured P2P systems can achieve excellent scalability and reliability. The performance of the proposed scheme has been benchmarked against a super-peer overlay topology based on a mesh graph using the flooding with TTL value 7. The theoretical results showed that the Multi-hop Index Replication with PDG-based construction scheme yield a higher query success ratio, a reduced number of search messages, and a lower average hop-count delay. It would be interesting for our future work to investigate how the heterogeneity affects our proposal.

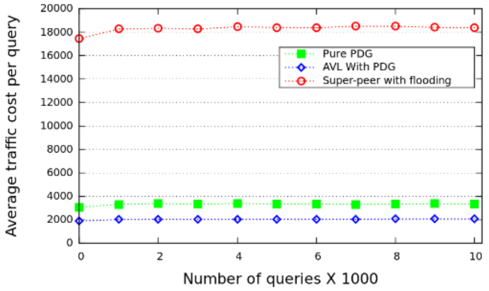


Figure 4. Comparison of number of broadcast search messages in mesh-based and PDG overlay networks

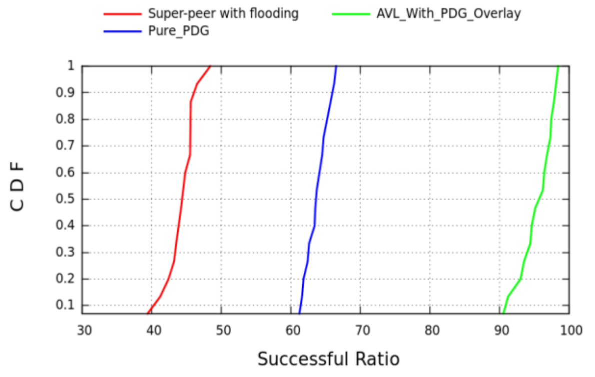


Figure 5. Comparison of query successful ratio in mesh-based and PDG overlay networks.

TABLE IV. AVERAGE VALUE COMPARISON

| Scheme | Volume of Traffic | Success Ratio |
|---|---|---|
| Mesh-Based | 100% | 45.6% |
| Pure-PDG | 18.15% | 62.5% |
| AVL-PDG | 11.09% | 96.6% |

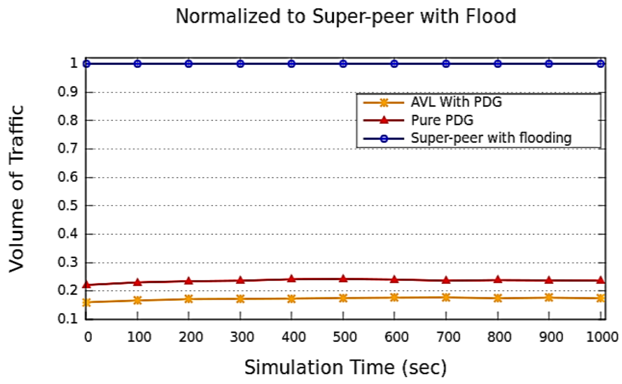


Figure 6. Normalized Network Traffic in mesh-based and PDG overlay networks.

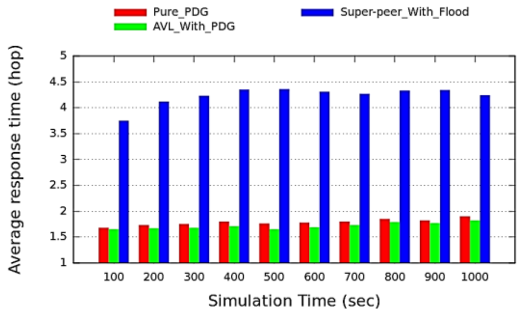


Figure 7. Average response time incurred in mesh-based and PDG overlay networks.

## References

[1] S. Banerjee, C. Kommareddy, K. Kar, S. Bhattacharjee, and S. Khuller, *"Construction of an Efficient Overlay Multicast Infrastructure for Real-Time Applications,"* Proc. IEEE INFOCOM '03, pp. 1521-1531, Mar. 2003.

[2] S. Sen and J. Wang, *"Analyzing Peer-to-Peer Traffic Across Large Networks,"* Proc. Internet Measurement Workshop, Nov. 2002.

[3] Y. Chu, S. Rao, and H. Zhang, *"A Case for End System Multicast,"* Proc. ACM SIGMETRICS'00, pp. 1-12, June 2000.

[4] E. Brosh and Y. Shavitt, *"Approximation and Heuristic Algorithms for Minimum Delay Application-Layer Multicast Trees,"* Proc. IEEE INFOCOM '04, Mar. 2004.

[5] M. Freedman and R. Morris, *"Tarzan: A Peer-to-Peer Anonymizing Network Layer,"* Proc. ACM Conference on Computer and Communications Security, Nov. 2002.

[6] S. Tewari and L. Kleinrock, *"Proportional replication in peer-to-peer networks,"* Proc. IEEE INFOCOM '06, pp. 1-11, Apr. 2006.

[7] W. Acosta and S. Chandra, *"Understanding the practical limits of the Gnutella p2p system: An analysis of query terms and object name distributions,"* in MMCN, 2008.

[8] B. Parhami and M. Rakov, *"Perfect Difference Networks and Related Interconnection Structures for Parallel and Distributed Systems,"* IEEE Transactions on Parallel and Distributed Systems, VOL. 16, NO. 8, pp. 714-724, Aug. 2005.

[9] B. Parhami and M. Rakov, *"Performance, Algorithmic, and Robustness Attributes of Perfect Difference Networks,"* IEEE Trans. Parallel and Distributed Systems, vol. 16, no. 8, pp. 725-736, Aug. 2005.

[10] I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H. Balakrishnan, *"Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications,"* IEEE/ACM Transactions on Networking (TON), VOL. 11, NO. 1, pp. 17-32, Feb. 2003.

[11] C. Gkantisdis, M. Mihail, and A. Saberi, *"Random walks in peer-to-peer networks,"* Proc. IEEE INFOCOM '04, pp. 7-11, Mar. 2004.

[12] The Gnutella v0.6 Protocol, http://rfc-gnutella. sourceforge.net/src/rfc-0_6-draft.html, May 2014

[13] BitTorrent, http://www.bittorrent.com/, May 2014.

[14] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/, May 2014

[15] P. Ganesan, Q. Sun, and H. Garcia-Molina, *"YAPPERS: A Peer-to-Peer Lookup Service over Arbitrary Topology,"* Proc. IEEE INFOCOM '03, pp. 1250-1260, Apr. 2003.

[16] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, *"A Scalable Content Addressable Network,"* Proc. ACM SIGCOMM '01, pp. 161-172, Oct. 2001.

[17] B.T. Loo, R. Huebsch, I. Stoica, and J.M. Hellerstein, *"The Case for a Hybrid p2p Search Infrastructure,"* Proc. Workshop Peer-to-Peer Systems (IPTPS '04), pp. 141-150, Feb. 2004.

[18] http://Iwayan.info/Research/PeerToPeer/Papers_Research/P2P_Architecture/it02012.pdf, May 2014

[19] F. Wang, Y. Xiongand, and J. Liu, *"mtreebone: A hybrid tree/mesh overlay for application-layer live video multicast."* In The 27th IEEE International Conference on Distributed Computing Systems (ICDCS'07), Toronto, Canada, June 2007.

[20] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, *"Making Gnutella-like P2P systems scalable,"* Proc. ACM SIGCOMM '03, pp. 407-418, Jan. 2003.

[21] S. Ioannidis, and P. Marbach, *"Absence of Evidence as Evidence of Absence: A Simple Mechanism for Scalable P2P Search,"* Proc. IEEE INFOCOM '09, pp. 576-584, Apr. 2009.

# Investigation of Inadequate Users in a Q&A Site Who Use Two or More Accounts for Submitting Questions and Manipulating Evaluations

Hiroki Matsumoto, Yasuhiko Watanabe, Ryo Nishimura, Yoshihiro Okada, Shin Yamanaka
Ryukoku University
Seta, Otsu, Shiga, Japan
Email: t14m086@mail.ryukoku.ac.jp, watanabe@rins.ryukoku.ac.jp,
r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp, t100450@mail.ryukoku.ac.jp

***Abstract*—Some users in a question and answer (Q&A) site use multiple user accounts and attempt to manipulate communications in the site. In order to detect these inadequate multiple account users precisely, it is important to investigate them from various points of view. In this paper, we investigate users suspected of manipulating evaluations of their answers by using two or more accounts for submitting many questions. The results of this study will give us a chance to investigate purposes and behaviors of inadequate multiple account users in a Q&A site.**

***Keywords–multiple account; Q&A site; evaluation manipulation; credibility.***

## I. Introduction

These days, many people use question and answer (Q&A) sites, where users share their information and knowledge. Q&A sites offer greater opportunities to users than search engines in the following points:

1) Users can submit questions in natural and expressive sentences, not keywords.
2) Users can submit ambiguous questions because other users give some supports to them.
3) Communications in Q&A sites are interactive. Users have chances to not only submit questions but give answers and, especially, join discussions.

As a result, Q&A sites are promising media. One of the essential factors in Q&A sites is anonymous submission. In most Q&A sites, user registrations are required for those who want to join the Q&A sites. However, registered users generally do not need to reveal their real names to submit messages (questions, problems, answers, comments, etc.). It is important to submit messages anonymously to a Q&A site. This is because anonymity gives users chances to submit messages without regard to shame and reputation. However, some users abuse the anonymity and attempt to manipulate communications in a Q&A site. For example, some users use multiple user accounts and submit messages to a Q&A site inadequately. Manipulated communications discourage other submitters, keep users from retrieving good communication records, and decrease the credibility of the Q&A site. As a result, it is important to detect users suspected of using multiple user accounts and manipulating communications in a Q&A site. In this case, identity tracing based on user accounts is not effective because inadequate users are likely to hide their true identity to avoid detection. A possible solution is authorship identification based on analyzing stylistic features of messages. In recent years, a large number of studies have been made on authorship identification [1] [2] [3] [4] [5], however, few researchers addressed the identification issues of authors suspected of using multiple user accounts and manipulating communications in a Q&A site. To solve this problem, we proposed methods of detecting

- multiple account users suspected of submitting questions and their answers repeatedly [6], and
- multiple account users suspected of submitting many answers to the same question repeatedly [7].

However, little is known about the purposes and methods of inadequate multiple account users. As a result, it is important to investigate these inadequate multiple account users from various points of view. One example is how many accounts these inadequate users use for submitting questions and manipulating their evaluations. It is natural for them to use two or more user accounts for submitting questions and manipulating evaluations of their answers. It is because they do not want to draw attention to themselves. As a result, in this paper, we investigate users suspected of using two or more user accounts for submitting many questions and manipulating evaluations of their answers.

Finally, we should notice that it is difficult to verify the credibility of our investigation. This is because there is no reliable information about users who used multiple user accounts and manipulated communications in Q&A sites. In order to discuss the credibility of our investigation, we show the results of our investigation in detail. The results of this study will give us a chance to investigate purposes and behaviors of users who use multiple user accounts and intend to manipulate communications in a Q&A site.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we describe Yahoo! chiebukuro for an example of Q&A sites. In Section IV, we describe how inadequate users use multiple user accounts in Q&A sites. In Section V, we show how we detect users suspected of using two or more accounts for submitting questions and manipulating evaluations of their answers. In

TABLE I. THE NUMBERS OF QUESTIONERS AND THEIR QUESTIONS AND ANSWERERS AND THEIR ANSWERS IN YAHOO! CHIEBUKURO (FROM APRIL/2004 TO OCTOBER/2005).

| | number of questioners | number of questions | number of answerers | number of answers |
|---|---|---|---|---|
| the data of Yahoo! chiebukuro | 165,064 | 3,116,009 | 183,242 | 13,477,785 |

Section VI, we show the result of the investigation. Finally, in Section VII, we present our conclusions.

## II. RELATED WORKS

One of the essential factors of the Internet is anonymity. Joinson discussed the anonymity on the Internet from various points of view [8]. These days, many users abuse the anonymity. Take Sybil attack for example. In a Sybil attack, the attacker intends to gain large influence on a peer-to-peer (P2P) network by creating and using a large number of pseudonymous identities [9] [10]. Sybil attacks are cheap and efficient way to gain large influence on P2P networks [11]. Similarly, in human online communities, such as, web-based bulletin boards, chat rooms, and blog comment forms, many users are thought to use multiple user accounts inadequately and submit inadequate messages, such as, deceptive opinion spams. In recent years, a large number of studies have been made on authorship identification [1] [2] [3] [4] [5], however, few researchers addressed the identification issues of authors suspected of using multiple user accounts and manipulating communications in the Internet. One of the difficulties of this problem is that we did not have sufficient number of examples of inadequate multiple account users. To solve this problem, some researchers tried to extract inadequate submissions by using heuristic methods based on text similarities and ranking results [12] [13]. On the other hand, the authors of [14] pointed that these heuristic methods were insufficient to detect inadequate submissions precisely, and showed they could detect inadequate submissions precisely when they used large number of examples of inadequate submissions. However, they obtained examples of inadequate submissions by using Amazon Mechanical Turk [15]. The examples of inadequate submissions created by workers in Amazon Mechanical Turk have the following problems.

- Little is known about the purposes and methods of inadequate submissions. As a result, it is possible that their instructions to workers in Amazon Mechanical Turk were insufficient.
- There are unreliable workers in Amazon Mechanical Turk [16].

As a result, it is important to obtain inadequate submissions from the Internet. To solve this problem, we proposed methods of detecting inadequate multiple account users and their submissions [6] [7]. However, as mentioned, little is known about the purposes and methods of inadequate multiple account users. As a result, it is important to investigate these inadequate multiple account users and their inadequate submissions from various points of view.

## III. YAHOO! CHIEBUKURO

Yahoo! chiebukuro is one of the most popular community sites in Japan. Users of Yahoo! chiebukuro submit their questions and answers in the next way.

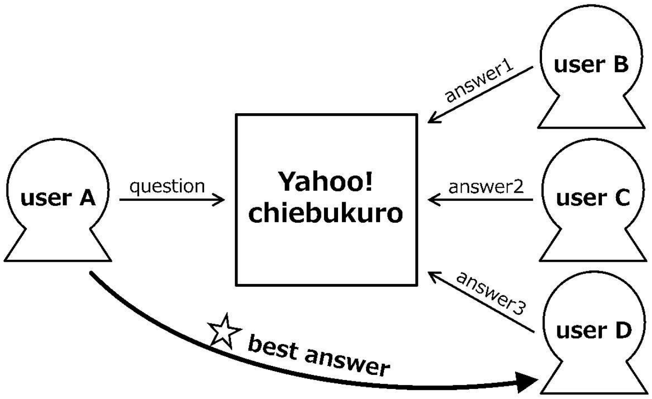


Figure 1. An example of how to use Yahoo! chiebukuro.

- user registrations are required for those who want to join Yahoo! chiebukuro.
- users do not need to reveal their real names to submit their questions and answers.
- each user can submit his/her answer only one time to one question.
- The period limit for accepting answers is one week. However, questioners can stop accepting answers before the time limits.
- After the time limits, questions with no answers are removed and cannot be referable. On the other hand, questions with answers can be referable.
- each questioner is requested to determine which answer to his/her question is best and give a *best answer* label to it.

Figure 1 shows that user A submitted one question to Yahoo! chiebukuro and three users, user B, user C, and user D answered the question, and then, user A selected user D's answer as a best answer. In this study, we used the data of Yahoo! chiebukuro for observation and examination. Chiebukuro means pearls of wisdom. The data of Yahoo! chiebukuro was published by Yahoo! JAPAN via National Institute of Informatics in 2007 [17]. This data consists of about 3.11 million questions and 13.47 million answers which were posted on Yahoo! chiebukuro from April/2004 to October/2005. In the data, each question has at least one answer because questions with no answers were removed. In order to avoid identifying individuals, user accounts were replaced with unique ID numbers. By using these ID numbers, we can trace any user's questions and answers in the data. Table I shows

- the numbers of questioners and their questions in the data, and
- the numbers of answerers and their answers in the data.

In Table I, the number of questioners is the number of users who submitted one or more questions to Yahoo! chiebukuro from April/2004 to October/2005. Also, the number of answerers is the number of users who submitted one or more answers to Yahoo! chiebukuro from April/2004 to October/2005.

## IV. SUBMISSIONS BY USING MULTIPLE USER ACCOUNTS

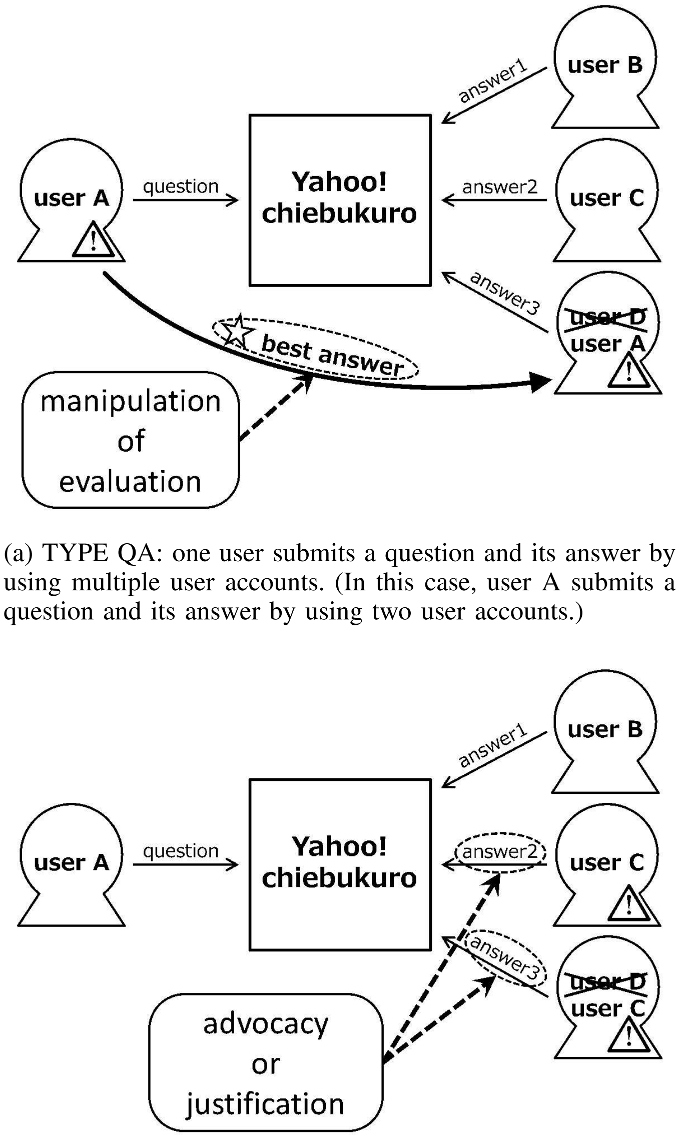
There are many reasons why users in a Q&A site use multiple user accounts. First, we discuss a proper reason. In Yahoo! chiebukuro, users do not need to reveal their real names to submit their questions and answers. However, their submissions are traceable because their user accounts are attached to them. Because of this traceability, we can collect any user's submissions and some of them include clues of identifying individuals. As a result, to avoid identifying individuals, it is reasonable and proper that users change their user accounts or use multiple user accounts. However, the following types of submissions by using multiple user accounts are neither reasonable nor proper.

**TYPE QA** One user submits a question and its answer by using multiple user accounts (Figure 2 (a)).
We think that the user intended to manipulate the submission evaluation. For example, in Yahoo! chiebukuro, each questioner is requested to determine which answer is best and give a *best answer* label to it. These evaluations encourage answerers to submit new answers and increase the credibility of the Q&A site. We think, the user repeated this type of submissions because he/she wanted to get many best answer labels and be seen as a good answerer.

**TYPE AA** One user submits two or more answers to the same question by using multiple user accounts (Figure 2 (b)).
We think that the user intended to dominate or disrupt communications in the Q&A site. To be more precise, the user intended to

- manipulate communications by advocating or justifying his/her opinions, or
- disrupt communications by submitting two or more inappropriate messages.

TYPE AA submissions are more similar to Sybil attacks in P2P networks than TYPE QA submissions. The more answers inadequate users submit by using multiple user accounts, the easier they manipulate or disrupt communications in a Q&A site.



(a) TYPE QA: one user submits a question and its answer by using multiple user accounts. (In this case, user A submits a question and its answer by using two user accounts.)

(b) TYPE AA: one user submits two or more answers to the same question by using multiple user accounts. (In this case, user C submits two answers by using two user accounts.)

Figure 2. Two types of inadequate submissions: TYPE QA and TYPE AA.

These two types are not all types of inadequate submissions. However, these kinds of submissions seriously disrupt communications in a Q&A site. Especially, TYPE QA submissions are serious because users can manipulate evaluations of messages by repeating TYPE QA submissions. Manipulated evaluations discourage other submitters, keep users from retrieving good communication records, and decrease the credibility of the Q&A site. Furthermore, we think we cannot use knowledge and countermeasures obtained in studies of Sybil attacks in P2P networks because TYPE QA submissions are different from Sybil attacks. In a Sybil attack, the more pseudonymous identities the attacker uses, the easier he/she gain large influence on a P2P network. On the other hand, in a TYPE QA submission, the inadequate user can get a best answer label by using only two user accounts. To solve this problem, we proposed methods of detecting multiple account users suspected of repeating TYPE QA submissions [6]. However, little is known about the purposes and methods of inadequate multiple account users. As a result, it is important to investigate these inadequate multiple account users from various points of view. For example, it is important to investigate how many accounts these inadequate users use for submitting questions and manipulating evaluations of their answers. Inadequate multiple account users can be classified into two types:

- inadequate users each of whom use one user account for submitting questions and manipulating evaluations of his/her answers, and
- inadequate users each of whom use two or more user accounts for submitting questions and manipulating evaluations of his/her answers.

In this study, we investigate the latter type of users, in other words, users suspected of using two or more user accounts for submitting questions and manipulating evaluations of their answers.

## V. Detection of users suspected of using two or more accounts for submitting questions and manipulating evaluations

Suppose that one user intends to manipulate evaluations of his/her answers, submitted by using user account $a$, and repeats TYPE QA submissions by using two user accounts, $q_1$ and $q_2$. In this case, it is expected that we observe the following abnormal submissions:

- user $a$ submits too many answers to questions submitted by user $q_1$ and $q_2$,
- user $q_1$ and $q_2$ receive too many answers from user $a$, and
- user $q_1$ and $q_2$ give too many best answer labels to user $a$'s answers.

In order to detect users who intend to manipulate evaluations of their answers and submit many questions by using two or more user accounts, we propose a method which consist of the following two steps:

1) We first detect user pairs of questioner and answerer, which are suspected of repeating TYPE QA submissions, as shown in Figure 2, by using three hypotheses: Hypothesis QA1, QA2, and QA3.
2) We detect users who are answerers in two or more user pairs detected by using Hypothesis QA1, QA2, and QA3.

Hypothesis QA1, QA2, and QA3 are as follows:

*a) Hypothesis QA1:* If user $a$ did not submit abnormally too many answers to user $q$'s questions, we would expect that user $a$ submitted at most $N_{QA1}(q,a)$ answers to user $q$'s questions.

$$N_{QA1}(q,a) = P_{QA1}(q) \times ans(a) \tag{1}$$

where $ans(a)$ is the total number of answers submitted by user $a$ and $P_{QA1}(q)$ is the probability that an user selects one question randomly and the question is one of user $q$'s questions. Because each user of Yahoo! chiebukuro can submit his/her answer only one time to one question, $P_{QA1}(q)$ is

$$P_{QA1}(q) = \frac{qst(q)}{N_{qst}} \tag{2}$$

where $qst(q)$ is the number of questions submitted by user $q$ and, as shown in Table I, $N_{qst}$ is the total number of questions in the data of Yahoo! chiebukuro. If this hypothesis is rejected by an one-sided binomial test, we determine that user $a$ submitted abnormally too many answers to user $q$'s questions.

The binomial test is an exact test of the statistical significance of deviations from a theoretically expected binomial distribution of observations into two categories [18]. There are two types of binomial tests: one sided binomial test or two sided binomial test. When the critical area of a distribution is one-sided, in other words, it is either greater than or less than a certain value, but not both, only the one-sided binomial test is generally applicable. In this study, the distribution area is one-sided, we use the one-sided binomial test.

*b) Hypothesis QA2:* If user $q$ did not receive abnormally too many answers from user $a$, we would expect that user $q$ received at most $N_{QA2}(q,a)$ answers from user $a$.

$$N_{QA2}(q,a) = P_{QA2}(a) \times qst(q) \tag{3}$$

where $qst(q)$ is the total number of questions submitted by user $q$ and $P_{QA2}(a)$ is the probability that an user received one answer from user $a$ when user $a$ selected one question randomly and answered it. Because each user of Yahoo! chiebukuro can submit his/her answer only one time to one question, $P_{QA2}(a)$ is

$$P_{QA2}(a) = \frac{ans(a)}{N_{qst}} \tag{4}$$

where $ans(a)$ is the number of answers submitted by user $a$ and, as shown in Table I, $N_{qst}$ is the total number of questions in the data of Yahoo! chiebukuro. If this hypothesis is rejected by an one-sided binomial test, we determine that user $q$ received abnormally too many answers from user $a$.

*c) Hypothesis QA3:* If user $q$ did not give abnormally too many best answer labels to user $a$'s answers, we would expect that user $q$ gave at most $N_{QA3}(q,a)$ best answer labels to user $a$'s answers.

$$N_{QA3}(q,a) = P_{QA3}(q) \times f_{QA}(q,a) \tag{5}$$

where $f_{QA}(q,a)$ is the number of answers submitted by user $q$ to user $q$'s questions, and $P_{QA3}(a)$ is the best answer ratio of user $a$.

$$P_{QA3}(a) = \frac{bestans(a)}{ans(a)} \tag{6}$$

where $ans(a)$ is the number of answers submitted by user $a$ and $bestans(a)$ is the number of best answers in user $a$'s answers. However, if user $j$ satisfies one of the following conditions:

- all user $a$'s answers were selected as best answers, in other words,
$$ans(a) = bestans(a) \tag{7}$$
- Hypothesis QA3aux, the auxiliary hypothesis for Hypothesis QA3, is rejected, in other words, it is considered that user $a$ received too many best answer labels,

we set $P_{QA3}(a)$ as follows:

$$P_{QA3}(a) = \frac{N_{bestans}}{N_{ans}} = \frac{N_{qst}}{N_{ans}} \tag{8}$$

ISBN: 978-1-61208-349-0

TABLE II. THE DETECTION RESULT OF USERS SUSPECTED OF USING TWO OR MORE USER ACCOUNTS FOR SUBMITTING QUESTIONS AND MANIPULATING EVALUATIONS OF THEIR ANSWERS

| significance levels for<br>QA1, QA2, QA3, and QA3aux | $UP_{BT}$ | $UP_{two\_or\_more}$ | $A_{BT}$ | $A_{two\_or\_more}$ |
|---|---|---|---|---|
| 0.00005 | 814 | 329 | 581 | 96 |
| 0.00001 | 603 | 222 | 450 | 69 |
| 0.000005 | 537 | 188 | 408 | 59 |
| 0.000001 | 424 | 135 | 333 | 44 |
| 0.0000005 | 407 | 129 | 319 | 41 |
| 0.0000001 | 337 | 104 | 266 | 33 |
| 0.00000005 | 325 | 101 | 257 | 33 |
| 0.00000001 | 278 | 86 | 220 | 28 |

$UP_{BT}$ is the number of user pairs which are detected by binomial tests based on Hypothesis QA1, QA2, QA3, and QA3aux. $UP_{two\_or\_more}$ is the number of user pairs the answerers of which were found in two or more user pairs detected by binomial tests. $A_{BT}$ is the number of answerers which are found in user pairs detected by binomial tests based on Hypothesis QA1, QA2, QA3, and QA3aux. $A_{two\_or\_more}$ is the number of answerers which are found in two or more user pairs detected by binomial tests based on Hypothesis QA1, QA2, QA3, and QA3aux.

where $N_{bestans}$ is the total number of best answers. $N_{bestans}$ is equal to $N_{qst}$ because each question has one best answer. If this hypothesis is rejected by one-sided binomial test, we determined that user $q$ gave abnormally too many best answer labels to user $a$'s answers.

*d) Hypothesis QA3aux:* If user $a$ did not receive abnormally too many best answer labels, we would expect that user $a$ received at most $N_{QA3aux}(a)$ best answer labels.

$$N_{QA3aux}(a) = P_{QA3aux} \times ans(a) \tag{9}$$

where $P_{QA3aux}$ is the average best answer ratio.

$$P_{QA3aux} = \frac{N_{bestans}}{N_{ans}} = \frac{N_{qst}}{N_{ans}} \tag{10}$$

If this hypothesis is rejected by one-sided binomial test, we consider that user $a$ received abnormally too many best answer labels.

## VI. Result of the Investigation

To evaluate our method, we conducted the detection of users suspected of using two or more user accounts for submitting many questions and repeating TYPE QA submissions, and manipulating evaluations of their answers in a Q&A site. In this experiment, the target users were all submitters in the data of Yahoo! chiebukuro. As shown in Table I, the numbers of the target questioners and answerers in the data of Yahoo! chiebukuro are 165,064 and 183,242, respectively.

In our method, we varied the significance levels for Hypotheses QA1, QA2, QA3, and QA3aux from 0.00005 to 0.00000001. They were extremely low because we intend to detect extreme abnormal submissions. Table II shows the results of this experiment.

As shown in Table II, 59 users were detected when the significance level was 0.000005. We should notice that 28 users of them were detected when the significance level was 0.00000001. It shows that many users were detected although the significance level was extremely low. As we expected, there are many users suspected of repeating TYPE QA submissions and manipulating evaluations of their answers by using two or more user accounts for submitting questions.

We checked questions and answers submitted by the detected user pairs and found that some other questioners were criticized for their unfair best answer selections. For example, user 233650 was criticized that he/she selected user 678451's answers as best answers repeatedly and unfairly. After criticized for his/her unfair best answer selection, user 233650 stopped submitting any questions to Yahoo! chiebukuro. Our method is useful for detecting these suspicious users. Furthermore, if we detect and take care of these suspicious users, we can avoid unnecessary frictions between users.

## VII. Conclusion

In this study, we investigated users suspected of using two or more user accounts for submitting questions and manipulating evaluations of their answers. We first discuss reasons why users in a Q&A site use multiple user accounts. We think many users use multiple user accounts reasonably and properly, however, some users use multiple user accounts improperly. For example, there seem to be users who use two or more user accounts for submitting questions and manipulating evaluations of their answers. However, little is known about the purposes and methods of these inadequate users. As a result, in order to investigate these inadequate users, we proposed a detecting method based on binomial test in this paper. Then, we applied our method to the data of Yahoo! chiebukuro, and found that many users suspected of using two or more user accounts for submitting questions and manipulating evaluations of their answers although the significance level was extremely low. We intend to use the results of this study for further investigation of purposes and behaviors of inadequate multiple account users in Q&A sites. For example, it is important to investigate which and how many categories inadequate multiple account users tried to manipulate evaluations. Also, it is important to analyze what inadequate multiple account users mentioned in their questions and answers. Furthermore, we intend to avoid unnecessary frictions between users in Q&A sites by detecting and taking care of these inadequate users.

## References

[1] O. de Vel, A. Anderson, M. Corney, and G. Mohay, "Mining e-mail content for author identification forensics," SIGMOD Rec., vol. 30, no. 4, Dec. 2001, pp. 55–64. [Online]. Available: http://doi.acm.org/10.1145/604264.604272 [accessed: 2014-04-26]

[2] M. Koppel, S. Argamon, and A. R. Shimoni, “Automatically categorizing written texts by author gender,” Literary and Linguistic Computing, vol. 17, no. 4, Nov. 2002, pp. 401–412. [Online]. Available: http://dx.doi.org/10.1093/llc/17.4.401 [accessed: 2014-04-26]

[3] M. Corney, O. de Vel, A. Anderson, and G. Mohay, “Gender-preferential text mining of e-mail discourse,” in Proceedings of the 18th Annual Computer Security Applications Conference (ACSAC ’02), Dec. 2002, p. 282.

[4] S. Argamon, M. Šarić, and S. S. Stein, “Style mining of electronic messages for multiple authorship discrimination: First results,” in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’03), Aug. 2003, pp. 475–480. [Online]. Available: http://doi.acm.org/10.1145/956750.956805 [accessed: 2014-04-26]

[5] R. Zheng, J. Li, H. Chen, and Z. Huang, “A framework for authorship identification of online messages: Writing-style features and classification techniques,” Journal of the American Society for Information Science and Technology, vol. 57, no. 3, Feb. 2006, pp. 378–393. [Online]. Available: http://dx.doi.org/10.1002/asi.v57:3 [accessed: 2014-04-26]

[6] N. Ishikawa, Y. Watanabe, R. Nishimura, K. Umemoto, Y. Okada, and M. Murata, “Detection of users suspected of using multiple user accounts and manipulating evaluations in a community site,” in Proceedings of the 6th IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE’10), Aug. 2010, pp. 600–607.

[7] N. Ishikawa, K. Umemoto, R. Nishimura, Y. Watanabe, and Y. Okada, “Detection of users in a Q&A site who suspected of submitting multiple answers to a question by using multiple user accounts,” in Proceedings of the fourth International Conferences on Internet Technologies and Applications (ITA 11), Sep. 2011, pp. 236–244.

[8] A. N. Joinson, Understanding the Psychology of Internet Behaviour: Virtual Worlds, Real Lives. Palgrave Macmillan, Feb. 2003.

[9] J. R. Douceur, “The sybil attack,” in Proceedings of the First International Workshop on Peer-to-Peer Systems (IPTPS ’02), Mar. 2002, pp. 251–260. [Online]. Available: http://www.cs.rice.edu/Conferences/IPTPS02 [accessed: 2014-04-26]

[10] L. A. Cutillo, M. Manulis, and T. Strufe, “Security and privacy in online social networks,” in Handbook of Social Network Technologies and Applications, B. Furht, Ed. Springer, Nov. 2010, pp. 497–522.

[11] L. Wang and J. Kangasharju, “Real-world sybil attacks in bittorrent mainline dht,” in Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM 2012), Dec. 2012, pp. 826–832. [Online]. Available: http://dx.doi.org/10.1109/GLOCOM.2012.6503215 [accessed: 2014-04-26]

[12] N. Jindal and B. Liu, “Opinion spam and analysis,” in Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM ’08), Feb. 2008, pp. 219–230. [Online]. Available: http://doi.acm.org/10.1145/1341531.1341560 [accessed: 2014-04-26]

[13] G. Wu, D. Greene, B. Smyth, and P. Cunningham, “Distortion as a validation criterion in the identification of suspicious reviews,” in Proceedings of the First Workshop on Social Media Analytics (SOMA ’10), Jul. 2010, pp. 10–13. [Online]. Available: http://doi.acm.org/10.1145/1964858.1964860 [accessed: 2014-04-26]

[14] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT ’11) - Volume 1, Jun. 2011, pp. 309–319. [Online]. Available: http://dl.acm.org/citation.cfm?id=2002472.2002512 [accessed: 2014-04-26]

[15] “Amazon Mechanical Turk,” URL: http://www.mturk.com/ [accessed: 2014-05-12].

[16] C. Akkaya, A. Conrad, J. Wiebe, and R. Mihalcea, “Amazon mechanical turk for subjectivity word sense disambiguation,” in Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk (CSLDAMT ’10), Jun. 2010, pp. 195–203. [Online]. Available: http://dl.acm.org/citation.cfm?id=1866696.1866727 [accessed: 2014-04-26]

[17] “Distribution of “Yahoo! Chiebukuro” data,” URL: http://www.nii.ac.jp/cscenter/idr/yahoo/tdc/chiebukuro_e.html [accessed: 2014-05-12].

[18] M. Hollander and D. A. Wolfe, Nonparametric Statistical Methods, 2nd Edition. Wiley-Interscience, Jan. 1999.

# Design of Auto-Tuning PID Controller Methods Based on Genetic Algorithms for LR-PONs

Tamara Jiménez, Noemí Merayo, Juan C. Aguado, Ramón J. Durán, Ignacio de Miguel, Patricia Fernández, Rubén M. Lorenzo, Evaristo J. Abril

Optical Communications Group of the Department of Signal Theory, Communications and Telematic Engineering,
E.T.S.I. Telecommunication, University of Valladolid, Valladolid, Spain
e-mail: tamara.jimenez@tel.uva.es, noemer@tel.uva.es, jaguado@tel.uva.es, rduran@tel.uva.es,
ignacio.miguel@tel.uva.es, patfer@tel.uva.es, rublor@tel.uva.es, ejad@tel.uva.es

***Abstract*—In this paper, a new method to automate the tuning process of PID controllers is presented. The designed method, based on genetic algorithms, tunes PID systems that control the QoS requirements in Long-Reach PONs. This new tuning technique has been compared with the manual Ziegler-Nichols frequency response method. The simulation results have demonstrated that the new technique efficiently automates the tuning process, which leads to a reduction in the tuning time and a higher accuracy.**

***Keywords- Proportional-Integral-Derivative (PID); Passive Optical Network (PON); tuning process; Dynamic Bandwidth Allocation (DBA); Class of Service (CoS); Service Level Agreement (SLA); bandwidth guarantees; delay guarantees.***

## I. INTRODUCTION

Passive Optical Networks (PONs) and Long-Reach Passive Optical Networks (LR-PONs) are the most preferable networks infrastructures in the today's access deployment [1]. In fact, the number of PON subscribers in Asia Pacific remains 80 million subscribers by the end of 2012, whereas in America this number is 11 million and in Europe near 16 million users [2]. However, in Europe, 41.5 million households are expected to be biber access subscribers by means of PON infrastructures at end of 2017. On the other hand, current access networks have to deal with different kind of users which contract a Service Level Agreement (SLA) with a provider and different kind of Class of Services (CoS) with different priorities. Therefore, users are guaranteed some network requirements, typically related to a minimum bandwidth level or a maximum delay for high priority CoS. Consequently, it is highly necessary that Dynamic Bandwidth Allocation (DBA) algorithms cover one or both premises. Even more, it is quite suitable that algorithms comply with the network requirements by means of a real time and automatic readjustment. Some algorithms take into account both objectives in a very efficient way [3]-[6]. Hence, one typical way to guarantee bandwidth or delay bounds to different priority subscribers is using fixed weights assigned to each ONU according to its SLA. Hence, ONUs that belong to a higher priority SLA, are assigned a larger weight, so they are given more bandwidth [7][8]. However, fixed factors do not adapt the PON performance to different traffic patterns or network conditions, so if service providers do not properly adjust the initial weights, the network should automatically evolve to the requirements established by the service provider. Thus, it is essential that the network becomes independent of the initial weights or conditions. Therefore, algorithms based on Proportional-Integral-Derivative (PID) controllers are able to robustly and efficiently manage the allocated bandwidth to comply with different guaranteed bandwidth levels [5] or maximum delay requirements [6]. Indeed, these algorithms based on PID controllers have demonstrated better performance than other existing algorithms that control these network parameters (bandwidth, delay) in PON networks [5][6]. This kind of controllers is extensively used due to its simple structure, robustness and good performance [9][10]. In connection with this type of control, PIDs require a tuning process in order to achieve a reliable response according to the established objectives. However, in contrast to previous existing algorithms based on PIDs to control network parameters in PON infrastructures [5][6], which use the well-known Ziegler-Nichols frequency response method, we propose to tune the PID controller using a Genetic Algorithm (GA). Although, the Ziegler-Nichols frequency response method is a very widespread technique, it is a manual method based on experiments. Thus, this manual nature may convert it into a very time-consuming and tedious technique. Contrary, the use of GA allows an automatic and fine tuning process, with less tuning time and better accuracy than manual techniques.

Therefore, in this paper a GA is developed to tune the PIDs of the previous developed algorithms [5] and [6] to automatize the tuning process and to improve their performance when controlling the bandwidth and the delay network parameters. The rest of this paper is organized as follows. Section II describes the genetic algorithms developed to auto-tune PID controllers to control network parameters in LR-PONs. Section III presents the simulation results and the discussion. Section IV addresses the conclusions of the paper.

## II. GENETIC ALGORITHMS TO AUTO-TUNING PID CONTROLLERS TO CONTROL NETWOK PARAMETERS

### *A. Genetic algorithm to tune a PID to control the bandwith allocation in PONs*

A PID controller is designed to keep the value of a variable close to a desired value [11]. Therefore, the PID calculates the error, defined as the difference between the current value of the variable and its reference value. According to this committed error ($e[n]$), it calculates the

ISBN: 978-1-61208-349-0

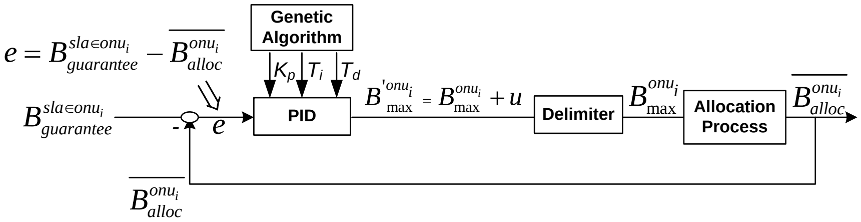

$e = B_{guarantee}^{sla \in onu_i} - \overline{B_{alloc}^{onu_i}}$
Genetic Algorithm
$K_p$
$T_i$
$T_d$
$B_{guarantee}^{sla \in onu_i}$
-
e
PID
$B'^{onu_i}_{\max} = B^{onu_i}_{\max} + u$
Delimiter
$B^{onu_i}_{\max}$
Allocation Process
$\overline{B_{alloc}^{onu_i}}$
$\overline{B_{alloc}^{onu_i}}$

Figure 1. Block diagram of the proposed PID controler tuned with a GA to ensure guranteed bandwidth levels to differnet priority profiles.

control signal $u[n]$ following (1), which is the equation that models the PID in the discrete time. It is composed by three terms, the proportional term manages the current error, the integral one regards the accumulation of past errors, and the derivative term makes a prediction of future errors [11].

$$u[n] = K_p \cdot e[n] + K_p \cdot \frac{T_{sample}}{T_i} \sum_{m=0}^{n} e[m] + K_p \cdot \frac{T_d}{T_{sample}} (e[n] - e[n-1]) \quad (1)$$

The first proposed algorithm, called Genetic Algorithm Service level agreement PID (GA-SPID), keeps the mean allocated bandwidth of each ONU ($\overline{B_{alloc}^{onu_i}}$) close to its guaranteed bandwidth, which depends on its contracted SLA ($B_{guarantee}^{sla \in onu_i}$). GA-SPID assigns bandwidth to each Optical Network Unit (ONU) at every cycle, $B_{alloc}^{onu_i}$, using a polling policy with a limited scheme, defined as $B_{alloc}^{onu_i} = Minimum\left\{B_{demand}^{onu_i}, B_{\max}^{onu_i}\right\}$, where $B_{demand}^{onu_i}$ is the bandwidth demanded by ONU $i$ in one cycle (in bytes). To include the control of the PID in the bandwidth allocation ($B_{alloc}^{onu_i}$), the term $B_{\max}^{onu_i}$ is updated by adding the control signal. Therefore, the maximum bandwidth allowed to each ONU, $B_{\max}^{onu_i}$, is dynamically updated depending on the committed error, that is, the difference between the mean allocated bandwidth and the required bandwidth $e[n] = B_{guarantee}^{sla \in onu_i} - \overline{B_{alloc}^{onu_i}}$. In case that one ONU demands less bandwidth than its guarantee value, the PID only offers its demand because the remaining bandwidth up to its guarantee level will be unused by the ONU and the EPON performance could become inefficient. Finally, the system includes a delimiter which reduces the maximums proportionally to the ones calculated by the controller, to fit in the maximum cycle time of the Ethernet PON (EPON) standard (2 ms). Fig. 1 shows the block diagram of the proposed PID for the bandwidth assignment (GA-SPID).

On the other hand, the parameters $K_p$, $T_i$ and $T_d$ of (1) have to be tuned so that the control system will be stable and meet the established objectives. Among the existing techniques, the frequency response method proposed by Ziegler-Nichols [9][11] has become an easy and very high spread technique, especially when a mathematical model is not available, as in our system. It gives simple and experimental rules by only considering the proportional response ($T_i = \infty, T_d = 0$) and then the gain is increased until the process begins to oscillate. When this happens, the gain is defined as the ultimate gain ($K_u$) and the oscilation period is defined as the ultimate period ($T_u$). With both variables, it is possible to obtain $K_p$, $T_i$ and $T_d$ following a simple relation [11]. This method has been previously used in [5] to tune the PID for the bandwidth allocation process in a PON network. However, it can be noticed that it is a manual technique that sometimes may become a time-consuming and laborious method if the selected values are quite far of the suitable ones. Therefore, we propose an automatic method based on genetic algorithms to select the tuning parameters in a very efficient way. This method, as well as the Ziegler-Nichols method, is an offline tuning method, carried out just before the PON activates the PID which controls the network parameters. Indeed, genetic algorithms, which are efficient searching techniques used to optmize parameters and processes, have been included in the tuning process of PID controllers in many fields. In the literature, there are some proposals in different chemical and industrial applications [12][13]. In the Telecom field, it could be found one genetic algorithm that tunes a PID controller with the aim to improve the network utilization [14].

In order to design the novel tuning method, the main steps of genetic algorithms were followed. The first step regards the definition of the chromosome. In our system, each chromosome consists of the three tuning parameters ($K_p$, $T_i$, $T_d$) coded in a binary chain (16 bits per parameter), since this type of codification improves the efficiency of the genetic algorithm for this application [12]-[14] (Fig. 2).

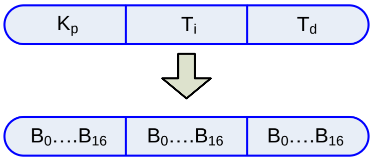


Figure 2. Appearance of a chormosome of the tuning parameters

After that, a random initial population is created. For the specific application of tuning a PID, the number of individuals that compose the population may become critical due to the strong dependence between the population size and the tuning time. Indeed, a low population size leads to a fast evolution of the algorithm towards the optimum tuning values. The next step consists of evaluating the fitness of each member of the population. Since the objective of the algorithm is to minimize the error between the desired output and the one obtained, to calculate the fitness of one specific individual, an objective function based on the committed error of the PID using this individual ($K_p$, $T_i$, $T_d$), during $m$ iterations of the PID has

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-349-0

been used. Specifically, the objective function for one individual is defined according to (2), where $N_{onus}$ is the number of ONUs in the PON and $e_i[m]$ is the error committed by ONU $i$ in the $m$ iteration of the PID.

$$F = \frac{1}{m} \cdot \frac{1}{N_{onus}} \sum_{m} \sum_{i=0}^{N_{onus}} \left| e_i[m] \right| \qquad (2)$$

Therefore, those individuals with the lowest error (which are the fittest) have a high probability to be selected for the next iteration of the genetic algorithm, in which a new generation is created. Once the fitness evaluation of each individual is finished, the genetic algorithm checks the stop criterion. If the criterion is not satisfied, the algorithm repeats the process with the next generation. To generate a new population, the genetic algorithm selects individuals according to its fitness and it applies the crossover and mutation operators. Furthermore, in GA-SPID we have considered elitism, which means that the fittest individual in each generation is retained unchanged, so that the best solution is not lost. If the stop criterion is satisfied, then the best individual of the population is used to tune the PID. Specifically, the stop criterion selected for our proposal is to reach a maximum number of generations, which allow us to establish a fixed duration of the tuning process. A flow diagram with the steps of the genetic algorithm is shown in Fig. 3.

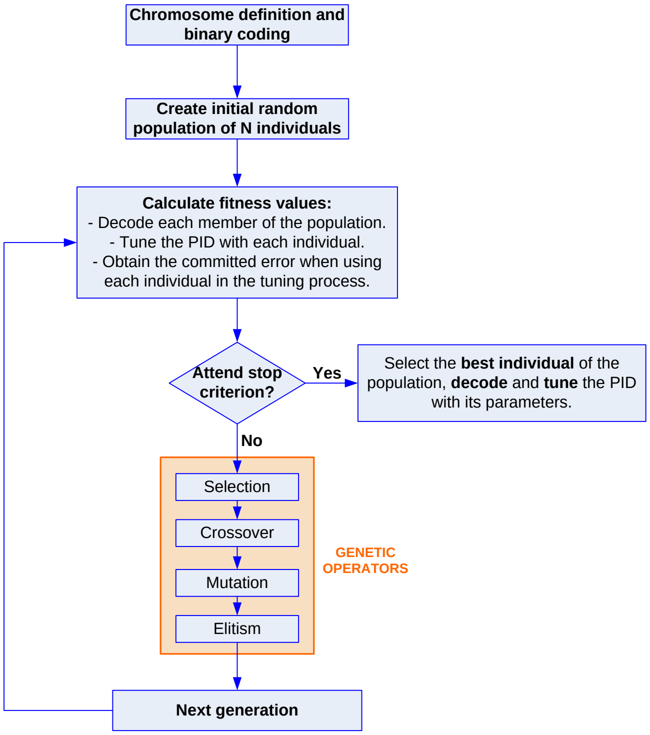

Chromosome definition and binary coding
Create initial random population of N individuals
Calculate fitness values:
- Decode each member of the population.
- Tune the PID with each individual.
- Obtain the committed error when using each individual in the tuning process.
Attend stop criterion?
Yes
Select the best individual of the population, decode and tune the PID with its parameters.
No
Selection
Crossover
Mutation
Elitism
GENETIC OPERATORS
Next generation

Figure 3. Flow diagram of the genetic algorithm used to tune the PID controller

### B. *Genetic algorithm to tune a PID to control the mean packet delay of priority services*

The algorithm Genetic Algorithm Delay aware Service level agreement PID (GA-DaSPID) is able to control the maximum delay of different priority services by modifying the maximum bandwidth of each SLA using a simple P controller, which is a simplification of the PID controller which only considers the proportional term of (1). The block diagram of the algorithm is shown in Fig. 4. Since this algorithm controls the mean packet delay, the reference value is the maximum permitted delay stipulated by the service provider for the $j$ classes of service with restrictive delay depending on the contracted $k$ SLA $(R_{P_j}^{sla_k})$. The term under control is the instantaneous mean packet delay for each $j$ class of service of each $k$ SLA $(\overline{r_{P_j}^{sla_k}[n]})$. In order to calculate the instantaneous error ($e[n]$) of one ONU which belongs to the SLA $k$, it is necessary to carry out the sum of every individual committed error in each service $j$ in order to guarantee the delay restrictions to each $j$ class of service, that is, $e = \sum_j (R_{P_j}^{sla_k} - \overline{r_{P_j}^{sla_k}[n]})$. On the other hand, to calculate the control signal $u[n]$, (1) is applied with only the proportional term, since a P controller has demonstrated the best performance for this concrete application [6]. To obtain the new maximum permitted bandwidth ($B_{\max}^{onu_i}$), the control signal $u[n]$ is subtracted from the previous maximum permitted bandwidth. In this way, if for example the mean packet delays of all $j$ services of $ONU_i$ $(\overline{r_{P_j}^{sla_k}[n]})$ are higher than their maximum packet delay $(R_{P_j}^{sla_k})$, the error becomes negative, so the algorithm increments the maximum permitted bandwidth to allow $ONU_i$ to decrease its mean packet delay so that it can comply with the delay restrictions. In contrast, if the mean packet delays of all $j$ services of $ONU_i$ $(\overline{r_{P_j}^{sla_k}[n]})$ are lower than their maximum packet delay $(R_{P_j}^{sla_k})$ the error becomes positive and the P controller reduces its maximum allocated bandwidth. As in the previous algorithm, the designed P controller is equipped with a delimiter (Fig. 4).

As in GA-SPID, a new method to efficiently tune the P controller by using a genetic algorithm is proposed. Since the controller only consists of the proportional term, the tuning parameters are reduced to $K_p$. Consequently, the chromosome is a binary code of 16 bits that represents only this parameter ($K_p$). Furthermore, the steps of the genetic algorithm are the same as those represented in the flow diagram of Fig. 3. Finally, the objective function of the algorithm is also (2).

## III. RESULTS AND DISCUSSIONS

### A. *Simulation scenario of the LR-PON*

We have designed a LR-EPON network with 16 ONUs and one user connected to each ONU using OPNET Modeler v.16 [15]. The upstream and downstream transmission rates are 1 Gbit/s whereas the transmission rate from users to each ONU is 100 Mbit/s. The distance between ONUs and the Optical Line Terminal (OLT) is 100 km. The maximum cycle time according to the EPON

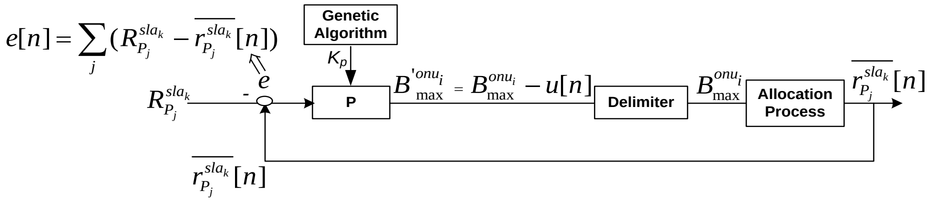

$e[n]=\sum_j (R_{P_j}^{sla_k} - \overline{r_{P_j}^{sla_k}}[n])$
Genetic Algorithm
$K_p$
$e$
$R_{P_j}^{sla_k}$
-
P
$B'^{onu_i}_{\max} = B^{onu_i}_{\max} - u[n]$
Delimiter
$B^{onu_i}_{\max}$
Allocation Process
$\overline{r_{P_j}^{sla_k}}[n]$
$\overline{r_{P_j}^{sla_k}}[n]$

Figure 4. Block diagram of the proposed P controller tuned with a GA to control delay requirements in priority services in PONs

standard is 2 ms [16]. Traffic follows a Pareto distribution with a Hurst parameter H equal to 0.8, with variable packet length between 64 and 1500 bytes, plus the 38 bytes of the packet headers It has been assumed symmetry in the traffic load of every ONU, as in [1][3]-[8]. Furthermore, GA-SPID considers three SLAs ($SLA_0$, $SLA_1$, $SLA_2$) with their corresponding guaranteed bandwidth levels of 100 Mbps, 75 Mbps and 50 Mbps, to be controlled by the PID. Regarding GA-DaSPID, it takes into account three services, $P_0$ for the highest priority traffic (interactive), $P_1$ for the medium priority traffic (responsively) and $P_2$ for the non-critical traffic (best-effort). $P_0$ assumes the 20% of the total network load, and $P_1$ and $P_2$ the 40% of the total load, as in [6]. Furthermore, it considers three SLAs ($SLA_0$, $SLA_1$, $SLA_2$), so each delay-sensitive service ($P_0$, $P_1$) is set a different bound delay depending on the priority of the profile. In fact, Table I summarizes the delay bounds considered in GA-DaSPID, which are the same that those proposed by other algorithms [4][6].

TABLE I. DELAY BOUNDS FOR EACH CLASS OF SERVICE AND SERVICE LEVEL AGREEMENT CONSIDERED IN GA-DASPID

| ***Class of Service*** | ***Delay bound value*** | ***Applications*** |
|---|---|---|
| $P_0$ | 1.5 ms | VoIP, videoconference, interactive games, Telnet |
| $P_1$ | $SLA_0$: 5 ms | Voice Messaging<br>Web-browsing HTML<br>E-mail<br>Transaction services |
| | $SLA_1$: 20 ms | |
| | $SLA_2$: 60 ms | |
| $P_2$ | Not limited | Bulk data |

The parameters related to the execution of the genetic algorithm for both algorithms are specified in Table II. These parameters have been selected by running previous simulations, and choosing those parameters which allows both, a good performance and a short tuning time. To justify the selection of these parameters, Fig. 5 represents the mean committed error (in bits) of the best individual in each generation when considering different population sizes and number of iterations in GA-SPID. As it can be observed, the error is reduced as the number of generations increases for every combination of population size and number of iterations. However, the worst performance is achieved for a population of 15 individuals and a number of iterations equal to 2. For the remaining combinations, the results are similar. Thus, a population of 20 individuals with 2 iterations is selected, since it leads to the lowest tuning time.

TABLE II. MOST IMPORTANT PARAMETERS OF THE GENETIC ALGORITHM IN GA-SPID AND GA-DASPID

| ***Parameters of the genetic algorithm*** | ***GA-SPID*** | ***GA-DaSPID*** |
|---|---|---|
| Selection method | Roulette wheel | Roulette wheel |
| Threshold of tuning parameters | (0,5] | (0,5] |
| Cross probability | 0.9 | 0.9 |
| Mutation probability | 0.01 | 0.01 |
| Elitism | yes | yes |
| Population size | 20 individuals | 15 individuals |
| Stop criteria | 10 generations | 10 generations |
| Iterations of the PID to update fitnesss | 2 iterations | 5 iterations |

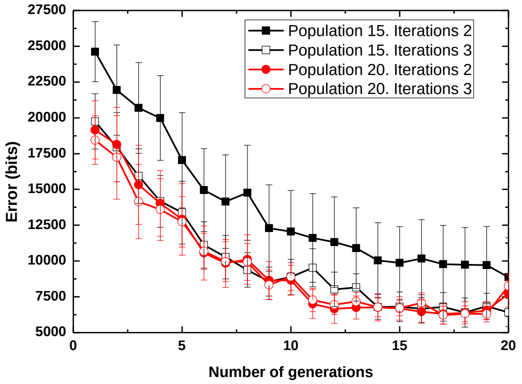


Figure 5. Evolution of the mean committed error of the best individual of each generation when considering different population sizes and number of iterations of the PID in GA-SPID.

A similar analysis has been carried out in the algorithm GA-DaSPID. Thus, Fig 6 represents the committed error (in seconds) of the best individual in each generation when considering different population sizes and number of iterations. It can be seen that the best performance is obtained with a population of 15 individuals and a number of iterations of the PID equal to 5.

Finally, in both algorithms the number of generations of the stop criterion is fixed to 10, because high number of generations increases the tuning time, but the reduction of the error (as it can be observed in Fig. 5 and Fig. 6) is not remarkably.

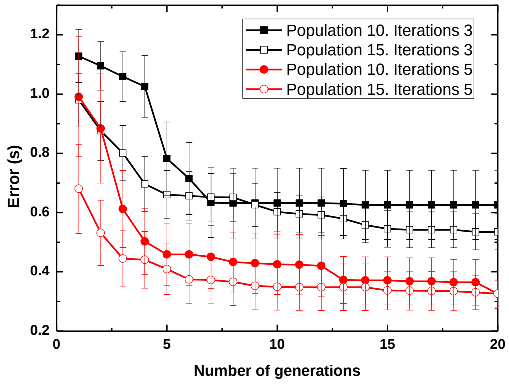


Figure 6. Evolution of the mean committed error of the best individual of each generation when considering different population sizes and number of iterations of the PID in GA-DaSPID.

## B. Simulation study of GA-SPID

The objective of GA-SPID is to guarantee minimum bandwidth levels to different priority profiles using a PID controller. In contrast to SPID [5], designed for the same purpose but tuned with the Ziegler-Nichols method, GA-SPID incorporates a genetic algorithm to automatically tune the PID. In order to emphasize the importance of a correct tuning process of the PID controller, Fig. 7 and Fig. 8 show the real time evolution of $B_{\max}^{onu_i}$ and the mean value of $B_{alloc}^{onu_i}$ for the SLA$_1$ profile, respectively, considering different values for the tuning parameters. In particular, we compare the genetic algorithm solution, with the Ziegler-Nichols solution used in [5] and a random configuration of the tuning parameters. As it can be observed in Fig. 7, the genetic algorithm obtains more stability in the maximum permitted bandwidth than the Ziegler-Nichols or the random solution. Besides, Fig. 8 demonstrates the same performance for the evolution of the mean allocated bandwidth. In fact, the genetic algorithm achieves a more stable response than the other two tuning methods when approaching to the guaranteed bandwidth of SLA$_1$ profile (75 Mbps). Therefore, the importance of optimizing the tuning process to design a reliable PID can be stated by observing the bad performance of the random tuning parameter in both graphs.

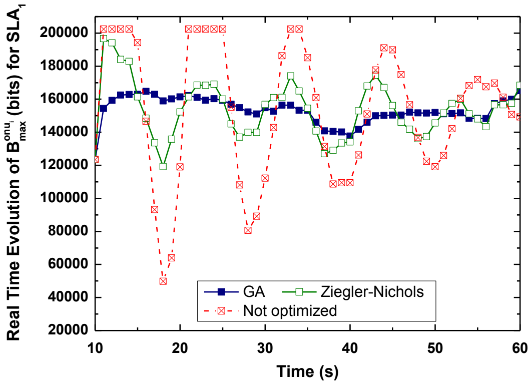


Figure 7. Real time evolution of the maximum permitted bandwidth for the SLA$_1$ profile considering different values for the tuning parameters

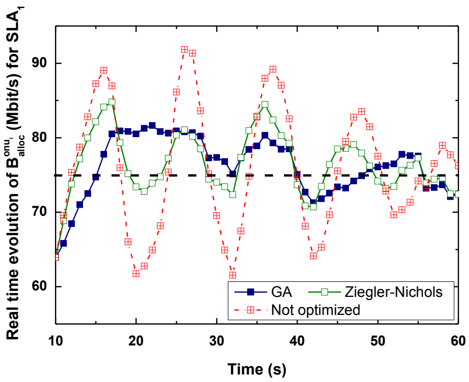


Figure 8. Real time evolution of the mean allocated bandwidth for the SLA$_1$ profile considering different values for the tuning parameters

On the other hand, the Ziegler-Nichols method is completely manual, based on visual oscillations of the controlled variable for different values of $K_p$. Therefore, if the chosen values are quite far from the suitable ones, the tuning process may become very slow. Once a value of $K_p$ is selected, the performance when it tunes the PID during a regular interval have to be observed. In case fluctuations at the end of the interval keep high, another $K_p$ value is necessary. In contrast, if fluctuations are low and kept inside a maximum and a minimum threshold, the selected value can be considered as a good $K_u$ and it can be used to tune the PID. Therefore, to compare the tuning process time of the genetic algorithm and the Ziegler-Nichols method, we propose to automate this last method. This way, we consider a random initial value of $K_p$ (between (0,5], as in the genetic algorithm) and its performance is observed during 300 seconds (a good interval to observe a more or less stable response). If fluctuations of the mean allocated bandwidth keep over the 10% above and below of the guarantee bandwidth (that is, the desired value for the variable under control), $K_p$ moves to another value in steps of 0.1. Once the $K_p$ value reaches the higher value of its interval (in this case 5), the following $K_p$ values are obtained from the random initial value in descending steps of 0.1 until the end of the lower threshold of the interval (in this case 0). The selected value of the step affects the tuning time. In fact, if a higher precision is needed, the step of 0.1 can be smaller, but it implies a higher tuning time. In contrast, if the step value increases, it could be difficult to achieve a good tuning process. On the other hand, when the oscillations are within the margin of 10%, the tuning process is finished. As an example, Fig. 9 represents the evolution of the mean allocated bandwidth of the SLA$_2$ profile when the initial value of $K_p$ is set to 2.7. Moreover, in blue and referred to the axis on the right, the variation of the $K_p$ values is represented. As it can be observed in the graph, for this initial random value of $K_p$ the tuning process last over 10000 s, since this value is far from the range of optimal $K_p$ values. Obviously, if the initial random value is near that range, the tuning process ends more quickly.

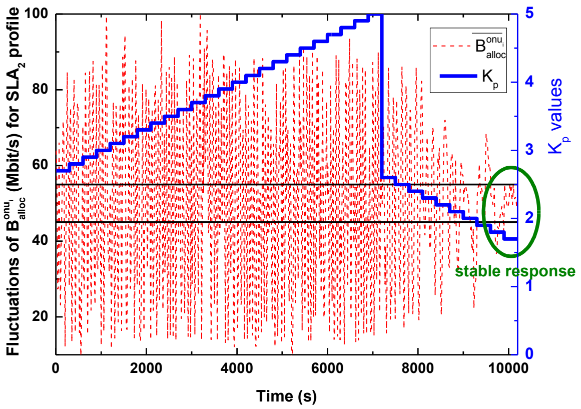


Figure 9. Tuning time of the Ziegler-Nichols method for the SLA$_2$ profile for a initial $K_p$ of 2.7

On the contrary, the tuning time in GA-SPID is lower. Indeed, the tuning time in this algorithm is given by (3), where $N$ is the population size of the genetic algorithm, $m$ is the number of iterations of the PID in which each individual is tested to obtain its fitness, $T_{sample}$ is the sample time used by the PID controller to obtain each error and $N_{Gen}$ is the number of generations of the stop criterion.

$$T_{tuning} = N \cdot m \cdot T_{sample} \cdot N_{Gen} \qquad (3)$$

Therefore, considering the values of Table II, and with a $T_{sample}$ equal to 1 s (which is a suitable value for GA-SPID), the maximum tuning time for GA-SPID is equal to 400 s. Hence, the great difference between both algorithms as regards the time to tune the PID can be noticed. Consequently, the main advantage of GA-SPID is related to the automation of the process, which involves a high reduction of the processing time. Moreover, thanks to the use of a genetic algorithm a more complete evaluation of the solution space is carried out, which leads to a more accurate tuning.

### C. Simulation study of GA-DaSPID

The main purpose of GA-DaSPID, as in DaSPID [6], is to control the mean packet delay of delay-sensitive applications taking into account client differentiation. This control is especially critical for high and medium network loads, when it is indispensable to efficiently assign the available resources so that all users comply with their network requirements. However, whereas in DaSPID the tuning process is made following the frequency response method of Ziegler-Nichols, GA-DaSPID uses a genetic algorithm. Therefore, in order to show the performance of GA-DaSPID, Fig. 10 and Fig. 11 represent the mean packet delay of P$_1$ for SLA$_1$ and SLA$_2$, respectively, for the highest network load, that is, ONUs transmitting at 100 Mbit/s. Only the performance of this class of service and these two user profiles is represented due to the lack of space. However, the performance of P$_0$ service for every profile and P$_1$ service for the SLA$_0$ profile is similar.

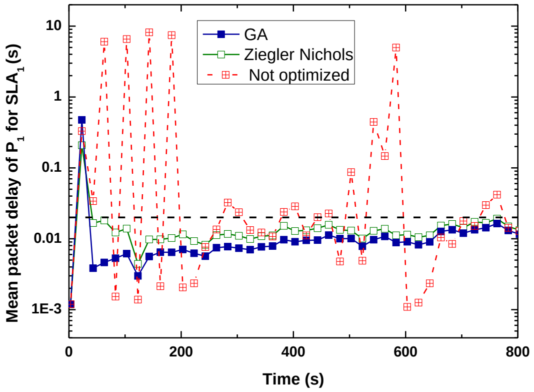


Figure 10. Real time evolution of the mean packet delay of P$_1$ for the SLA$_1$ profile considering different values for the tuning parameters

As it can be observed in Fig. 10 and Fig. 11, an optimum selection of the tuning parameters is essential to ensure a quick evolution of the mean packet delay under the limits specified for each profile. In fact, it can be appreciated that the not optimized solution is not able to keep the mean packet delay under the delay limits even in 800 s. In contrast, both Ziegler-Nichols and GA-DaSPID achieve this objective in less than 50 s. In this case, the differences between Ziegler-Nichols and GA-DaSPID are quite small, since the genetic algorithm has proposed a very similar solution to the Ziegler-Nichols method to tune the P controller.

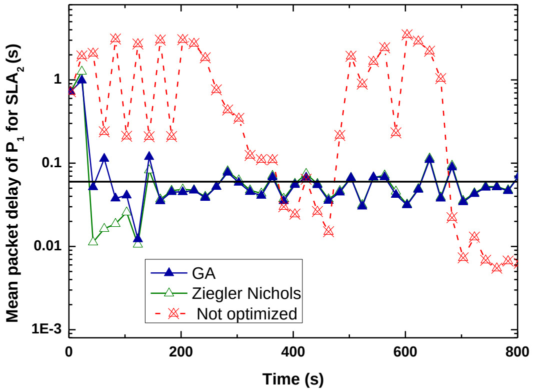


Figure 11. Real time evolution of the mean packet delay of P$_1$ for the SLA$_2$ profile considering different values for the tuning parameters

Regarding the comparison of the tuning time in both algorithms, Fig. 12 shows the results for the automated Ziegler-Nichols method to ensure delay limited bounds for SLA$_2$ profile when the initial $K_p$ value is equal to 2.1. In this case, the allowed range of fluctuations is a 30% above and below of the delay bound (60 ms). As it can be noticed, for this initial random value, the tuning time is higher than 11000 s. Obviously, if the initial $K_p$ value is near the optimal $K_p$ values, the tuning time will be lower.

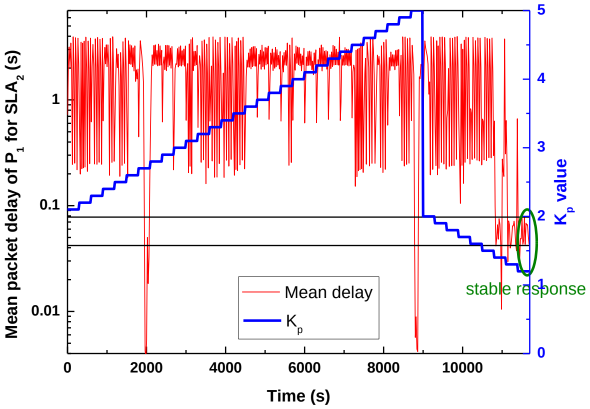


Figure 12. Tuning processing time of the Ziegler-Nichols method for the $SLA_2$ profile for a initial $K_p$ of 2.1

In contrast, for the GA-DaSPID algorithm, the tuning time is also given by (3). Thus, according to the parameters of Table II and with a $T_{sample}$ time of 10 s (which is the optimal value to control the delay [6]), the tuning time in GA-DaSPID is 7500 s. Therefore, the main advantage of GA-DaSPID is the efficient automation of the tuning process, which leads to a lower tuning time, as it happened in GA-SPID.

## IV. Conclusions

In this paper, the development of a genetic algorithm to tune PID controllers that have been designed to efficiently provide Quality of Service (QoS) in PON networks has been presented. In particular, one PID controller focuses on guaranteeing minimum bandwidth levels to different priority profiles, whereas the other one aims to provide delay requirements to different priority classes of service. In contrast to manual tuning techniques, such as the Ziegler-Nichols frequency response method, the genetic algorithm speeds up and automatically adapts the tuning process according to the stipulated objectives.

In order to demonstrate the benefits of this proposal over manual techniques, we have compared its performance with the Ziegler-Nichols frequency response tuning method. Simulation results have shown that the genetic algorithm efficiently automates the tuning process. Indeed, for the PID controller with bandwidth guarantees, the genetic algorithm allows a more stable response than Ziegler-Nichols for the mean allocated bandwidth and the maximum permitted bandwidth to every SLA. Regarding the P controller, which provides delay guarantees, the genetic algorithm achieves more stability of the mean packet delay of the high priority traffic ($P_0$, $P_1$). Furthermore, another important advantage of the genetic algorithm is a significant reduction of the tuning time, since, as it has been demonstrated, the Ziegler-Nichols method could become extremely time-consuming and quite tedious when calculating the tuning parameters. Consequently, the implementation of genetic algorithms to tune PID controllers provides a more accurate, efficient and fast performance than manual techniques, such as the Ziegler-Nichols frequency response method.

## References

[1] H. Song, B. W. Kim, and B. Mukherjee, “Long-Reach optical access networks: a survey of research challenges, demonstrations and bandwidth assignment mechanisms,” IEEE Communications Surveys & Tutorials, vol. 12, no. 1, pp. 112-122, 1st Quarter 2010.

[2] K. Ahl, FTTH Council Europe, “Creating a brighter future. A sustainable future enabled by fibre to the home,” [Online]. Available from:
www.scotland.gov.uk/Resource/0042/00425685.ppt 05.2014

[3] T. Jiménez et al., “Self-adapted algorithm to provide multi-profile bandwidth guarantees in PONs with symmetric and asymmetric traffic load,” Photonic Network Communication, vol. 24, no. 1, pp. 58-70, January 2012.

[4] N. Merayo et al., “EPON bandiwdth allocation algorithm based on automatic weight adaptation to provide client and service differentiation,” Photonic Network Communication, vol.17, no. 2, pp. 119-128, April 2009.

[5] T. Jiménez et al., “Implementation of a PID controller for the bandwidth assignment in Long-Reach PONs,” Journal of Optical Communications and Networking, vol. 4, no. 5, pp. 392-401, May 2012.

[6] T. Jiménez et al., “A PID-based algorithm to guarantee QoS delay requirements in LR-PONs,” Optical Switching and Networking, in press. D.O.I:http://dx.doi.org/10.1016/j.osn.2014.01.005

[7] C. H. Chang, N. M. Alvarez, P. Kourtessis, R. M. Lorenzo, and J. M. Senior, “Full-service MAC protocol for metro-reach GPONs,” Journal of Lightwave Technology vol. 28, no. 7, pp. 1016 – 1022, April 2010.

[8] N. Merayo et al., “Adaptive polling algorithm to provide subscriber differentiation in a long-reach EPON,” Photonic Network Communications, vol. 19, no.3, 257-264, June 2010.

[9] K. H. Ang, G. Chong, and Y. Li, “PID control system analysis, design and technology,” IEEE Transactions on Control Systems Technology, vol. 13, no. 4, pp. 559-576, July 2005.

[10] P. Cominos and N. Munro, “PID controllers: recent tuning methods and design to specification,” IEEE Control Theory and Applications, vol. 149, no. 1, pp. 46-53, Enero 2002.

[11] K. J. Aström and T. Hägglund, “Advanced PID control”. Research Triangle Park, NC: ISA-The Instrumentation, Systems, and Automation Society, 2006.

[12] D. A Wati and R. Hidayat, “Genetic algorithm-based PID parameters optimization for air heater temperature control,” 2013 International Conference on Robotics, Biomimetic, Intelligent Computational Systems (ROBIONETICS), Nov. 2013, pp. 30 – 34,

[13] M. J. Neath, A. K. Swain, U. K Madawala, and D. J. Thrimawithana, “An optimal PID controller for a bidirectional inductive power transfer system using multiobjective genetic algorithm,” IEEE Transactions on Power Electronics, vol. 29, no. 3, pp. 1523-1530, March 2014.

[14] C. K. Chen, H. H. Kuo, J. J Yan, and T. L. Liao, “GA-based PID active queue management control design for a class of TCP communication networks,” Expert Systems with Applications, vol. 36, no. 2 Part 1, pp. 1903-1913, March 2009.

[15] Opnet Modeler. [Online]. Available from: http://www.opnet.com 04.2014.

[16] IEEE 802.3ah Ethernet in the First File Task Force [Online]. Available from: http://www.ieee802.org/3/efm/public 04.2014.