INTERNET 2015

The Seventh International Conference on Evolving Internet

ISBN: 978-1-61208-435-0

October 11 - 16, 2015

St. Julians, Malta

INTERNET 2015 Editors

Eugen Borcoci, University “Politehnica” Bucharest, Romania
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Leszek Koszalka, Wroclaw University of Technology, Poland

INTERNET 2015

Forward

The Seventh International Conference on Evolving Internet (INTERNET 2015), held between October 11 - 16, 2015 - St. Julians, Malta, continued a series of events dealing 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.

The conference had the following tracks:

- Internet performance, monitoring and control
- Advanced Internet mechanisms

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

We take here the opportunity to warmly thank all the members of the INTERNET 2015 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to INTERNET 2015. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the INTERNET 2015 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope INTERNET 2015 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of the evolving Internet. We also hope that St. Julians, Malta provided a pleasant environment during the conference and everyone saved some time to enjoy the beauty of the city.

INTERNET 2015 Chairs

INTERNET Advisory Committee

Eugen Borcoci, University “Politehnica” Bucharest, Romania
Abdulrahman Yarali, Murray State 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
Tingyao Wu, Alcatel-Lucent/Bell Labs, USA

INTERNET 2015

Committee

INTERNET 2015 Advisory Committee

Eugen Borcoci, University “Politehnica” Bucharest, Romania
Abdulrahman Yarali, Murray State 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 2015 Industrial/Research Chairs

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

INTERNET 2015 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
Demetris Antoniades, University of Cyprus, Cyprus
Olivier Audouin, Alcatel-Lucent Bell Labs, France
Liz Bacon, University of Greenwich, UK
Jacques Bahi, University of Franche-Comté, France
Michael Bahr, Siemens AG, Germany
Andrzej Beben, Warsaw University of Technology, Poland
Nik Bessis, University of Derby, UK
Maumita Bhattacharya, Charles Sturt University - Albury, Australia
Kashif Bilal, COMSATS Institute of Information Technology, Pakistan
Bruno Bogaz Zarpelão, State University of Londrina (UEL), Brazil
Eugen Borcoci, University “Politehnica” Bucharest, Romania
Fernando Boronat Seguí, Universidad Politécnica De Valencia, Spain

Damian Bulira, Wroclaw University of Technology, Poland
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
Hsing-Chung Chen, Asia University, Taiwan
Shiping Chen, Sybase Inc., USA
Tzung-Shi Chen, National University of Tainan, Taiwan
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
Henry Hong-Ning Dai, Macau University of Science and Technology, China
Guillermo Diaz-Delgado, Universidad Autónoma de Querétaro (UAQ) / Queretaro State University (UAQ), Mexico
Ioanna Dionysiou, University of Nicosia, Cyprus
Martin Dobler, FH VORARLBERG - Dornbirn, Austria
Yingfei Dong, University of Hawaii, USA
Charalampos Doukas, CREATE-NET, Trento, Italy
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
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
Jing (Selena) He, Kennesaw State University, USA
Frans Henskens, University of Newcastle, Australia
Ching-Hsien Hsu, Chung Hua University, Taiwan
Wladyslaw Homenda, Warsaw University of Technology, Poland

Copyright Information

Table of Contents

Impact of Router Security and Address Translation Mechanisms on the Transmission Delay 1
Dominik Samociuk, Blazej Adamczyk, and Andrzej Chydzinski

A Simple Passive Method to Estimate RTT in High Bandwidth-Delay Networks 6
Iria Prieto, Mikel Izal, Eduardo Magana, and Daniel Morato

Congestion Avoidance TCP Improvements for Video Streaming 12
Dirceu Cavendish, Kazumi Kumazoe, Gaku Watanabe, Daiki Nobayashi, Takeshi Ikenaga, and Yuji Oie

AccessWeb Barometer - A Web Accessibility Evaluation and Analysis Platform 18
Ramiro Goncalves, Jose Martins, Frederico Branco, Jorge Pereira, Carlos Peixoto, and Tania Rocha

Session Management of a Variable Video Rate Streaming Session over Multi-Channel Networks 26
Rachel Behar, Yael Samet, and Ronit Nossenson

Optical Last Miles for Research and Education in the Czech Republic 34
Jan Radil, Lada Altmannova, Ondrej Havlis, Miloslav Hula, Stanislav Sima, and Josef Vojtech

FPGA Based TCP Session Features Extraction Utilizing Off-Chip Memories 38
Satoshi Fuchigami, Hajime Shimada, Yukiko Yamaguchi, and Hiroki Takakura

Reliable Assurance Protocols for Information Systems 43
Mahalingam Ramkumar and Somya Mohanty

An Investigation of a Factor That Affects the Usage of Unsounded Code Strings at the End of Japanese and English Tweets 50
Yasuhiko Watanabe, Kunihiro Nakajima, Haruka Morimoto, Ryo Nishimura, and Yoshihiro Okada

Impact of Router Security and Address Translation Mechanisms on the Transmission Delay

Dominik Samociuk, Blazej Adamczyk, Andrzej Chydzinski

Silesian University of Technology
Institute of Informatics, Poland
email: {dominik.samociuk; blazej.adamczyk; andrzej.chydzinski}@polsl.pl

***Abstract*—We study transmission delays on an IP router caused by security and address translation mechanisms. Using a high-precision device for traffic generation and measurements and a simulated topology of two hundred end systems, we test three mechanisms of the following types: Access Control Lists, Intrusion Prevention Systems and Network Address Translation. As we show, in some cases the delay changes only a little bit, when the mechanism is turned on. In most cases however, the impact of the mentioned mechanisms is non-negligible and may increase the delay ten times in worst-case scenarios.**

Keywords–Transmission delay; IP networks, Secure architecture; Router security.

I. INTRODUCTION

One of the most important performance characteristics of computer networks is delay – the time between sending and receiving data. Transmission delays are an inherent problem of communication quality, starting with intermittent conversations via Internet telephony, through the delays in the transmission of video, ending with targeting missiles on the battlefield.

In this paper, we investigate how popular security and address translation mechanisms affect delays in IP networks. In particular, we focus on mechanisms implemented with layer 4 addressing. In the experiments, we verify the impact of Access Control Lists (ACL), Intrusion Prevention Systems (IPS) and Network Address Translation (NAT) technology on the delay generated by the device on which these mechanisms are implemented. Of course, it is to be expected that additional packet processing introduces additional delay. However, it is impossible to say in advance if this is 1%, 100% or 10000% of extra delay. Therefore, the goal of this paper is to check what is the order of magnitude of the delay induced by the studied mechanisms.

ACL [1][2][3], introduced first in Unix systems for extensive control access to files, were further extended to network devices to use higher layers in order to verify the access rights to network resources. Universal ACL consists of the information about source and destination Internet Protocol (IP) address, network mask, and port/protocol of higher layers [4].

IPS is a method for detecting and blocking attacks in real time [5][6]. Two modes of operation of the IPS are available (see, e.g., [7]):

- ”Promiscuous” (Intrusion Detection System mode) – analyzes the traffic copy, which does not slow down the traffic, but cannot block attacks in real time.
- ”In-line” (IPS mode) – analyzes the original traffic, slowing it down. ”In-line” mode can, however, automatically block attacks in the real time.

A router with the IPS mechanism turned on operates in transparent mode [8]. This means that the system analyzes the traffic passing through the router as a transparent bridge, by analyzing the layers 2-7 and appropriately responding to the defined threats.

The paper is organized as follows. Section 3 contains an overview of the testbed prepared for the experiments. In Section 4 we present the results of the ACLs experiments. In Section 5 the influence of IPSs on delays is studied. Section 6 describes the impact of the NAT mechanisms on transmission delays. The paper is concluded in Section 7.

II. RELATED WORK

For now, research activities are polarized in the following directions. Firstly, studies on developing improved mechanisms, such as detecting and reducing redundancy in ACLs, [9], or classification, analysis and deleting conflicts in Intrusion Prevention and Detection Systems, [10], are carried out. In addition to the direction of improving actual features, there are studies on other architecture schemes, such as network virtualization and software-defined networks [11]. However, there are no research paper, validating security mechanism with high-precision hardware traffic generator.

III. TESTBED

The testing environment was built using a high-precision hardware traffic generator, which allows to generate artificial traffic with characteristics needed in the prepared test scenarios with full line rates, as well as measure and analyze the arriving packets with time precision of 20ns. Moreover, the generator enables simulation of a virtual topology composed of many interconnected devices.

Namely, the Ixia generator with XM2 casing was used [12]. XM2 dual-port casing provides a platform to build a topology-based Ixia’s test solutions. Working with the family of test applications, XM2 is the basis of a complete environment for testing the performance and operation of the network. The casing allows installation of different modules for traffic generation: up to 32 Gigabit Ethernet ports, up to sixteen 10G Ethernet ports, and a single Ethernet port 40G, 100G or a single dual-port 40/100G Ethernet. These modules provide the necessary processing to test the application layers 2-7, the signaling, voice and video transmission, etc.

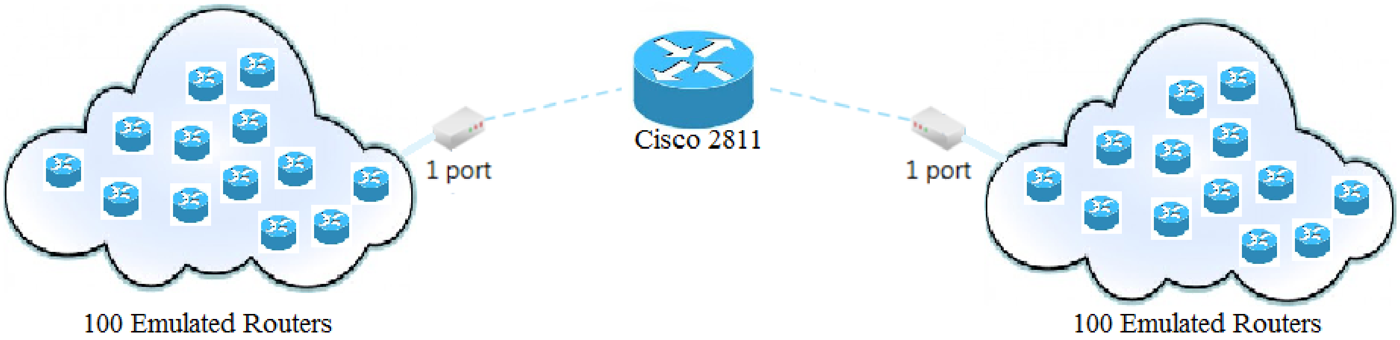


Figure 1. Virtual topology configuration.

The load module used in the tests, is an LSM1000XMVDC8-01 Gigabit Ethernet Load Module [13], offering full functionality for testing layers 2-7. Each port supports the generation and analysis of layers 2-3 with line rate, as well as high-performance emulation of routing and switching protocols. In order to monitor the traffic with high accuracy in real time, the device uses specialized programmable circuits. The load module used in this study was 8 ports (copper or fiber), operating in the range from 10Mbps to 1Gbps. Each port on the card has a separate RISC processor running Linux and a fully optimized stack for testing TCP/IP. This architecture provides the performance and flexibility in testing of routers, switches, broadband and wireless Internet access, access devices, web servers, video servers, gateways, firewalls, etc.

The Ixia's IxNetwork software is an application designed to test the performance and functionality of routers and switches [14]. IxNetwork works on separate modules and processors for each port. From the software perspective, each of them is a separate instance of Linux operating system. With this solution, each interface is tested independently, and the state of the corresponding instance is passed to the supervisor machine based on Microsoft Windows operating system.

IxNetwork software provides an easy-to-use graphical interface, which can be used to configure and run complex tests. Using IxNetwork tester, we can easily set up protocol variables and parameters specific to the needs of the device under test.

The specific testbed was chosen to emulate real traffic instead of just simulating it, which is usually the contemporary method nowadays. Ixia's hardware allow generate traffic with desired parameters, and then, with high-precision measured transmission delay generated by described security mechanisms. Devices chosen for tests have been selected to meet the specifications for possibility to configure discussed security mechanisms.

The configuration of the testing environment used in the experiments is depicted in Figure 1. The first topology is simulated on the input port, and the second topology on the output port of the Ixia's load module. Each topology consists of 100 devices. This is meant to simulate the connection between different pairs of addresses (Media Access Control- MAC, IP, etc.), transferred through the device under test. All the tests were carried on a single Cisco 2811 router (as in the middle of Figure 1), however due to similar architecture devices from the same class (access class devices for our studies) should generate comparable delays.

The traffic generated by the generator had the following parameters:

- direction of the flow, D, which was H or F (H meaning the alternating two-way traffic, i.e., half-duplex, F meaning the simultaneous two-way traffic, i.e., full-duplex),
- lack of optimization (Quality of Service (QoS) settings and IP Type of Service (ToS) Precedence),
- package size, S, in bytes,
- duration of the test, T, in seconds,
- load of the line, C, in percentage (e.g., 10% means that the percentage of transmission data including individual headers is 10% of the total capacity of the link).

The delay was measured from the time of completing the generation of the entire package to the last received bit on the receiver side (Last-In-Last-Out - LILO methodology). This is the default schema of time-stamping on Ixia devices.

Each test was repeated 1000 times. In the following sections, the resulting delays are presented in terms of the mean value and standard deviation based on the unloaded variance estimator.

IV. The impact of Access Control Lists on delay

The purpose of this set of tests was to verify the delay that is induced by the use and actions of ACLs in three scenarios:

- 100% of the traffic is proven through the ACL that allows traffic on the first rule,
- 100% of the traffic is proven through the ACL, in which the variable parameter is the number of traffic rules (all allowing traffic),
- 50% of the traffic is rejected by the ACL. The remaining traffic goes through a control list on the first rule and is checked whether the rejection affected the delay or not.

On the router, the standard and extended ACLs consisting of 1, 100, and 1000 dynamically generated entries were configured. The purpose of the test was to check what is the increase of delay when dealing with 1 and 1000 ACLs. An example of the extended ACL configuration with one entry (permitting traffic from 10.0.0.0/24 subnet to 10.0.1.0/24 subnet) is presented on Figure 2.

```
access-list 100 permit ip
    10.0.0.0 0.0.0.255 10.0.1.0 0.0.0.255
```

Figure 2. ACL configuration with one entry.

The tests of ACLs were carried out with disabled CEF (Cisco Express Forwarding) mechanism, [15].

Measurements were performed with the following parameters set:

- Direction: Full-duplex,
- Size: 64 B,
- Time: 30 seconds,
- Load: 10%,
- Number and type of checklists: a variable parameter.

TABLE I. RESULTS OF ACL TESTS.

The number and type of ACL rules	Delay [μs]
No ACL	147±1.23
1 – Standard list	150±1.89
100 – Standard list	270±2.5
100 – Extended list	310±3.3
1000 – Extended list	1500±32
2 streams – rejected + passed	150±1.93

The results of the experiments are presented in Table 1. As we can see, the implementation of ACLs may degrade significantly the observed delay. In particular, the usage of a single entry ACL had no significant effect on the delay, but a checklist of 100 entries increased the delay to 190% (standard list) and to 210% (extended list) of the original value. Exploiting ACL with 1000 entries increased the delay to 1000% of the base value.

It can be seen that the delay generated by ACLs increased approximately linearly with the number of rules.

V. Delays induced by the Intrusion Prevention System

These tests were performed to verify the IPS system overhead while scanning and detecting attacks in the traffic. The configuration of the router IPS is presented on Figure 3. The presented syntax, create ips rule, add signatures for basic vulnerabilities and enable it on the device. Prepared configuration allowed to check what is the transmission delay when traffic is passed through IPS mechanism with basic security rules.

The tests of IPS mechanisms were carried with enabled CEF mechanism. In the tests, the IPS mode was set to ”Inline”, which in addition to detection of attacks enables also preventing them in real time. The default thread signature was used [16], which provides a basic level of protection against a wide range of typical dangers.

Measurements were performed with the following parameters set:

- Direction: Full-duplex,
- Size: 64 B,
- Time: 30 seconds,
- Load: 10%.

```
ipips name sdm_ips_rule
ipips signature-category
category all
retired true
categoryios_ips basic
retired false
(...)
ipips signature-definition
signature 2004 0
status
retired false
enabled true
```

Figure 3. Configuration of the router IPS.

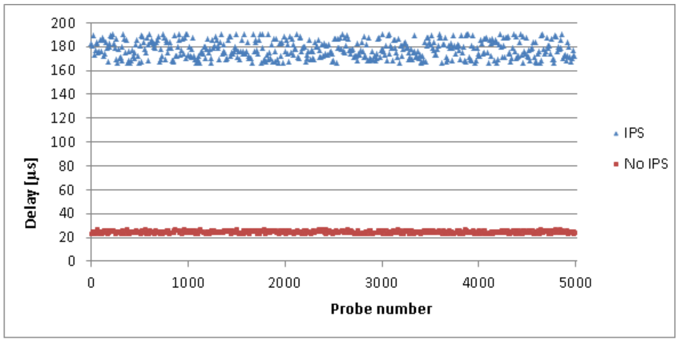


Figure 4. Distribution of probes in IPS tests.

Packet contain random data without any specific patterns to just pass-through IPS without raising any alarms. This payload type let measure actual IPS delay without false-positives with shorter delay, due to IPSs detection time.

The measured delay without IPS averaged at 26±1.2μs. Measured delay with IPS enabled averaged at 178±5μs. The values of the delay collected during the tests without and with the IPS mechanism are shown in Figure 4.

We can conclude that even a basic set of the IPS rules significantly increases the delay (700% of the initial value). Of course, the delay would be even greater for a larger number of signatures.

In the additional tests that were performed, with IPS enabled and operating in the ”Promiscuous mode”, the delay averaged at 28±0.6 μs. This shows that in the ”Promiscuous” mode, basically no additional delay is induced. (The addition of 2μs resulting from the need to copy the traffic flow on a different port is negligible). It must be remembered, however, that this mode does not provide protection in real time.

VI. Address translation impact on delay

Since the 90s, the IPv4 addressing space has been considered too small and the pool of addresses is still lowering. Creating the IPv6 standard solved the problem, but there are several issues that slow down migration to the new protocol [17][18]. Therefore, IPv6 is still not the most common method of preventing exhaustion of IPv4 addresses. Instead, local area networks use private addresses, which are translated into public

```
ip nat inside source static x.x.x.x
    x.x.x.x
```

Figure 5. Static NAT configuration.

```
ip nat pool xxxxxxxx PULA_NAT netmask
    255.255.255.0
ip access-list extended NAT
permit ip 10.0.0.0 0.0.0.255 any
ip nat inside source NAT pool lists
    PULA_NAT
```

Figure 6. Dynamic NAT configuration.

addresses using NAT method [19][20], when routed to the global network. RFC 1918 [21] describes the address class division and their pools due to the allocation of public and private parts. NAT concept was developed in three branches and implemented in three different ways in the network devices:

- Static Translation – one internal address is translated into one external address – no advantages associated with a reduction of usage of public IPv4 addresses.
- Dynamic Translation – some internal addresses are translated into several external addresses. The allocation is dynamically translated by the device.
- Port Address Translation (PAT) – several internal addresses are translated into one external address. Distinguishing between internal addresses is made by dynamic assignment of ports to them.

In this set of tests, the impact of NAT on network delays was verified. The following configuration was used.

- Static NAT configuration is presented on Figure 5 - where one internal IP address is translated to one external IP address.
- Dynamic NAT configuration is presented on Figure 6 - where internal IP addresses are translated to external IP addresses chosen from the specified pool.
- PAT configuration is presented on Figure 7 - where multiple internal IP addresses are translated to one external IP addresses.

The tests of NAT mechanisms were carried with disabled CEF mechanism. The measurements were performed with the following parameters set:

- Direction: Full-duplex,
- Size: 64 B,
- Time: 30 seconds,

```
ip nat pool xxxxxxxx PULA_NAT netmask
    255.255.255.255
ip access-list extended NAT
permit ip 10.0.0.0 0.0.0.255 any
ip nat inside source NAT pool
    PULA_NAT letter overload
```

Figure 7. PAT configuration.

TABLE II. RESULTS OF NAT TESTS.

Type of translation	**Delay [μs]**
No translation	147±2.6
Static NAT	151±2.8
Dynamic NAT	155±3
PAT	257±3.7

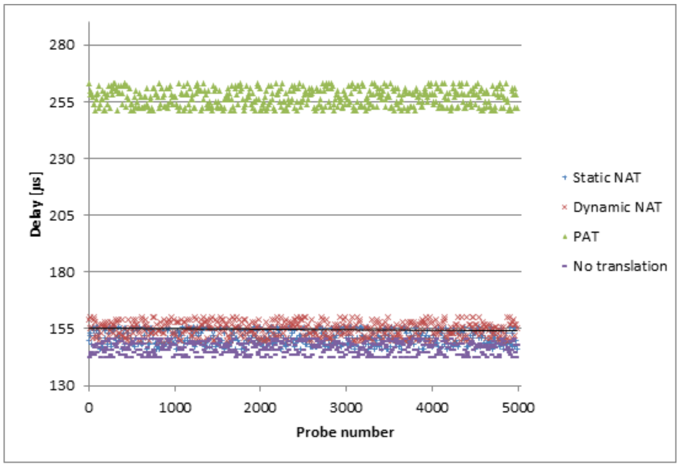


Figure 8. Distribution of probes in NAT tests.

- Load: 10%.

The results are presented in Table 2. As we can see, NAT in its static and dynamic versions does not introduce much overhead on the transmission delay. This has to be due to the simplicity of the operations that are executed and simple single cycles of the processor required for its implementation.

On the other hand, NAT with port translation (PAT) induces the delay of 175% of the original value. This is due to the need to use the layer 4 addressing of ports and analysis of data stored in the segment header.

Detailed test results are shown in Figure 8.

VII. Conclusion and Future Work

The studies conducted in the paper demonstrated the order of magnitude of additional delay induced by traffic filtering and security mechanisms. In the ACL case, the extra delay grows more or less linearly with the number of rules. For 100 rules the observed delay was twice as large as without ACL. For 1000 rules the delay increased 10 times. In the case of IPS set to in-line mode, the delay seven times larger than the original was observed. On the other hand, IPS in promiscuous mode had a negligible impact on the delay. Also the static and dynamic NAT had a minor impact of the delay. The PAT version, however, enlarged the delay by 75%.

As for the future work, the authors are working on a study of combined effects/mutual influence generated by described mechanisms. Also, an interesting continuation would be a study on the methodology of finding a secure topology design, while using as little overhead on the performance, as possible. In other words, the trade-off between the security and delay may be investigated. As long as we cannot allow for the degradation of security at the expense of increased performance, the solutions we are going to work on will focus on the migration to the new ways of creating network topology, inter alia, programmable networks.

REFERENCES

[1] J. Daly, A. X. Liu, and E. Torng, ”A difference resolution approach to compressing access control lists” in INFOCOM, 2013 Proceedings IEEE, 2013, pp. 2040 – 2048.

[2] R. Watson, ”A decade of OS access-control extensibility” in Communications of the ACM., v.56 n.2, 2013, pp. 52 – 63.

[3] L. Zhu, H. Mao, and H.Qin, ”A case study on Access Control Rules Design and Implementation of Firewall” in Proc. 8th International Conference on Wireless Communications, Networking and Mobile Computing, 2012, pp. 1 – 4.

[4] A. Sudarsan, A. Vasu, and D. Ganesh, ”Performance Evaluation of Data Structures in implementing Access Control Lists” in International Journal of Computer Networks and Security, vol. 24, issue 2, 2014, pp. 1303 – 1308.

[5] H. Ling – Fang, ”The Firewall Technology Study of Network Perimeter Security” in Proc. IEEE Asia-Pacific Services Computing Conference, 2012, pp. 410 – 413.

[6] M. Z. A. Aziz, M. Y. Ibrahim, A. M. Omar, R. A. Rahman, M. M. M. Zan, and M.I. Yusof, ”Performance analysis of application layer firewall” in Proc. IEEE Symposium on Wireless Technology and Applications (ISWTA), 2012, pp. 182 – 186.

[7] Z. Li, A. Das, and J. Zhou, ”Theoretical basis for intrusion detection” in IEEE workshop proceedings on information assurance and security, 2005, pp. 184 – 192.

[8] M. Gil-Jong, K. Yong-Min, K. Dong-Kook, and N. Bong-Nam, ”Network Intrusion Detection Using Statistical Probability Distribution” in Proc. Inter Conference: ICCSA(2), 2006, pp. 340 – 342.

[9] A. X. Liu, C.R. Meiners, and Y. Zhou, ”All-Match Based Complete Redundancy Removal for Packet Classifiers in TCAMs” in The 27th Conference on Computer Communications (INFOCOM), 2008, pp. 111 – 115.

[10] E. Al-Shaer, H. Hamed, R. Boutaba, and M. Hasan, ”Conflict classification and analysis of distributed firewall policies” in IEEE Journal on Selected Areas in Communications, vol.23, no.10, 2005, pp.2069 – 2084.

[11] R. Corin, M. Gerola, R. Riggio, F. De Pellegrini, and E. Salvadori, ”Network Virtualization and Beyond” in EWSDN, 2012, pp. 24 – 29.

[12] Ixia, http://ixiacom.com, [retrieved: July, 2015].

[13] Ixia Load Modules, http://www.ixiacom.com/sites/default/files/resources/datasheet/gigabit_ethernet_xmvdc_lan_services_modules.pdf, [retrieved: July, 2015].

[14] Ixia IxNetwork, http://www.ixiacom.com/products/ixnetwork, [retrieved: July, 2015].

[15] CEF Mechanism, http://www.cisco.com/c/en/us/td/docs/ios/12_2/switch/configuration/guide/fswtch_c/xcfcef.html, [retrieved: July, 2015].

[16] IOS IPS Routers, http://www.cisco.com/c/en/us/td/docs/security/security_management/cisco_security_manager/security_manager/4-1/user/guide/CSMUserGuide_wrapper/ipsios.html, [retrieved: July, 2015].

[17] T. Bilski, ”Network performance issues in IP transition phase” in Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference, 2010, pp. 39 – 44.

[18] K. Chakraborty, N. Dutta, and S. Biradar, ”Simulation of IPv4-to-IPv6 dual stack transition mechanism (DSTM) between IPv4 hosts in integrated IPv6/IPv4 network” in Computers and Devices for Communication, 2009. CODEC 2009. 4th International Conference, 2009, pp. 1 – 4.

[19] V. Krmicek, J. Vykopal, and R. Krejc, ”Netflow Based System for NAT Detection” in Co – Next Student Workshop09: Proc. International student workshop on emerging networking experiments and technologies, 2009, pp. 23 – 24.

[20] R. Li et. al., ”Passive NATted Hosts Detect Algorithm Based on Directed Acyclic Graph Support Vector Machine” in Proc. 2009 International Conference on Multimedia Information Networking and Security – Volume 02. MINES09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 474 – 477.

[21] RFC 1918, http://www.hjp.at/doc/rfc/rfc1918.html, [retrieved: July, 2015].

A Simple Passive Method to Estimate RTT in High Bandwidth-Delay Networks

Iria Prieto,
Mikel Izal,
Eduardo Magaña
and Daniel Morato

Public University of Navarre
Navarre, Spain
Email: iria.prieto, mikel.izal, eduardo.magana, daniel.morato @unavarra.com

***Abstract*—This paper presents a simple passive algorithm to estimate the Round-Trip-Time (RTT) of a TCP connection in high bandwidth-delay network scenarios. In these scenarios, a passive RTT estimator that can be used on captured packet traces is a useful tool for performance analysis. The algorithm is based on observing periodic RTT patterns in a TCP connection that is not filling its $bandwidth \times delay$ product. The results are compared to other passive methods such as the RTT of initial TCP handshake or TCP Timestamp option samples. The algorithm is shown to be effective and may be used in more scenarios than other methods thus, it provides a valuable tool to improve the amount of TCP connection whose RTT can be measured in a captured packet trace.**

Keywords–RTT; passive; network; traffic.

I. Introduction

Round-Trip-Time (RTT) is a key network performance metric. It is easy to measure using active probes (i.e., with Internet Control Message Protocol, ICMP) but it is not so simple to estimate by passive observation of traffic. There are situations where a passive RTT estimation from a captured packet trace is needed in the field of network and system analysis and network health check.

The RTT indicates the time that is taken to obtain a response from the other side of the communication, namely, the network latency. The latency can be used to identify short-life network problems. For instance, a RTT instantaneous peak may indicate a short period of congestion or suggest the existence of a network problem. Furthermore, the RTT of different network segments can be used to analyze if a performance issue is due to different parts of the network or even suggest it may be an application or server issue. RTT is a major factor in TCP, Transmission Control Protocol, connection and data transfer performance.

RTT is calculated as the elapsed time between one packet sent by one endpoint and the reception of a packet from the other endpoint that acknowledges the first packet. Any packet that can be guaranteed to have been sent by the other endpoint only after reception of the first packet may be used. There are several factors that can affect the measured time, like retransmissions or packet losses. Also, the other side may delay the response for protocol reasons (like TCP delayed ACK) or just because the response is not mandatory and the data flow in the other direction is being used as response. The later case can be defined as limited by the application.

All these factors make the passive estimating of RTT a non trivial task as it was exposed by Zhang [1]. It requires to take into account several conditions: disorders, retransmissions, lost packets, where the capture is located, etc.

Some RTT passive estimation methods have been proposed in the literature. The work of Strowes [2] is based on the observation of packets using TCP header timestamp option which is used by TCP protocol to generate RTT measures. For this samples to be present in a TCP connection the option TCP Timestamp option has to be activated, [3]. A TCP connection uses TCP options if both endpoints agree to use it at connection establishment negotiation. Another work which used Timestamp for RTT estimation is [4]. Both works show that the estimation solved some passive approach problems such as packet loss or capture point dependence. They also showed that Timestamp estimation is as good as the use of active ICMP probes. The main problem of TCP timestamp method is that there is still a very large fraction of TCP connections that do not use Timestamp option. As an example, our measurements at an university access link show only 22% of observed TCP connections successfully negotiated the use of TCP timestamp option. The test was performed on a trace of 1000 TCP connections captured during one work-day on Nov 2014.

Other passive methods work regardless of TCP timestamp option being used. For instance, the authors in [5] estimated the RTT value through the three-way handshake and the slow-start phase. The RTT provided by the three-way-handshake is not always accurate because it may be changed by middleboxes between client and server. These middleboxes may answer or initiate the connection on their own, resulting in lower RTTs. Besides, extracting parameters from time measures of TCP slow start phase is not an easy task. Other techniques relay on more complex mechanics. The authors in [6] associate a data segment with the ACK segment that triggered it. Other approaches try to measure RTT by mimicking changes in the sender's congestion window size [7]. It should be taken into account that these estimations are affected by packet losses, the TCP window scaling option and buggy TCP implementations.

The motivation for this work comes from the field of performance analysis of networks by means of captured packet traces. This is a valuable tool for troubleshooting and problem detection of large enterprise networks where packet traffic is captured at a vantage point to study network problems. This analysis is usually done by capturing traffic in advance and analyzing it a-posteriori in case some problem was reported. Thus, it is unfeasible to perform active measures of RTT. The large amount of traffic usually captured would make very difficult deciding, which endpoints to measure RTT in between. That is the reason why a passive RTT estimation tool is searched that can provide RTT values between the endpoints of every observed TCP connection. The scenarios of interest are high speed networks with middle to high traffic loads like those of datacenter or large enterprises.

The paper is organized as follows. First of all, the algorithm and configuration parameters are introduced. Section III describes the network scenario used to check the proposed algorithm. Section IV and V present the results and conclusions.

II. Proposed algorithms

The proposed algorithm provides an estimation of RTT by observing the behavior of TCP connections that are not filling its $bandwidth \times delay$ product. Note that a TCP connection data flow is limited by the flow control window which is advertised from each endpoint to the other.

Assuming an application, which always has data to send over TCP, if the RTT is high enough, TCP will be able to send the full permitted window of data and stop sending till it receives the confirmation for the first packet in the window. In that case data will be sent like and ON-OFF source with RTT period. If RTT is not so high, acknowledged packets will start arriving before the end of the window, resulting in more or less continuous data flow. The proposed RTT measuring method examines the TCP connection and infers the RTT from the observed ON-OFF pattern.

The idea is to provide an RTT estimation method for TCP connections requiring only passive capture of traffic. Even it may work just on some TCP connections depending on their advertised window and $bandwidht \times delay$ product, there are common scenarios where TCP window is small enough.

The algorithm evaluates if a given candidate RTT value could be the actual RTT seen by the TCP connection or it is an impossible value. The candidate is tested by using it as the time length interval to divide the connection time into slots and perform a simple check. If in any of the time slots more bytes have been sent by one endpoint than the advertised window, then the candidate is discarded. In fact, if a candidate is discarded it is clear that the actual RTT is lower than the candidate. If the candidate value passes the check it is an acceptable value for RTT. The algorithm searches for the smallest possible candidate value that can not be discarded as a valid RTT.

Defining the parameters,

- c: the total time that a connection lasts.
- t: the candidate time to be tested as possible value for the RTT.
- n: the number of t duration intervals on the connection $n = \lceil \frac{c}{t} \rceil$
- i: an interval being evaluated, $i \in 0..n$.
- B_i: Total bytes sent by the server in an interval, i.
- w_i: Maximum advertised window seen at interval i
- w_i^{max}: Maximum advertised windows seen up till interval i, $w_i^{max} = max\{w_k\} \quad \forall k \in 0..i$

The proposed algorithm consists in searching the smallest possible candidate t that does not fail the test. Several versions of the algorithm have been studied with different degrees of requirements. Depending on the test conditions different RTT estimators are generated named as RTT1, RTT2, RTT3.

1) RTT1: Candidate t is valid if in every interval, i, the total bytes sent is lower than the maximum window seen for each interval, equation 1
2) RTT2: Candidate t is valid if in every interval, i, the total bytes sent is lower than the maximum window seen for the whole TCP connection (discarding advertised window from TCP 3 way handshake packets SYN,SYN+ACK,ACK), equation 2
3) RTT3: Candidate t is valid if in every interval, i, the total bytes sent is lower than the maximum window seen up till that time in the TCP connection, equation 3

$$RTT1 = t \quad \forall i \in 0..n \quad B_i < w_i \tag{1}$$

$$RTT2 = t \quad \forall i \in 0..n \quad B_i < w_n^{max} \tag{2}$$

$$RTT3 = t \quad \forall i \in 0..n \quad B_i < w_i^{max} \tag{3}$$

In this work, the search for the smallest valid t is performed by initially choosing a candidate t that is clearly larger than the RTT and reducing t in a fixed amount δt every time the test is passed. When a value $t - \delta t$ fails the test, the previous t is declared the RTT estimation. The δt value used imposes the resolution of the estimator.

The algorithm is considered to finish when the candidate time fails the test. The value that will be shown as the final result is the previous one. The result of the algorithm is an upper limit from the theoretical RTT. In case that the first candidate time will not fail the test in the first iteration, it will not give any information since the real RTT could be higher than the tested one. In these cases, another candidate time could be chosen multiplying the first one by some value and the test would be restarted.

III. Validation scenario

The algorithm has been validated in the scenario of Figure 1. An emulated network of virtual machines is built with several client boxes in an emulated 10Mbps Ethernet. This virtual LAN, Local Area Network, is connected to a second virtual LAN through a routing virtual machine. In the second LAN there is a machine running a web server.

In the routing machine, the delay of packet forwarding is controlled using Netem tool [8]. The line speed of both virtual networks is 10Mbps.

The scenario is built with VirtualBox running on an Ubuntu 14.04 Linux PC. Client and router boxes are virtual boxes running Ubuntu 12.04. The host machine running the virtualization software acts also as the web server machine, Figure 1.

The scenario is configured for different RTTs by selecting the routing box forwarding delay, half RTT for each direction. The scenarios are configured for using total forwarding of

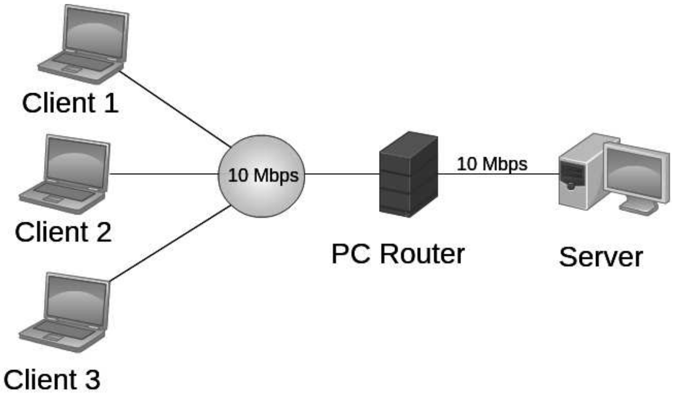


Figure 1. Emulated scenario of a network whose connections are limited by w/RTT.

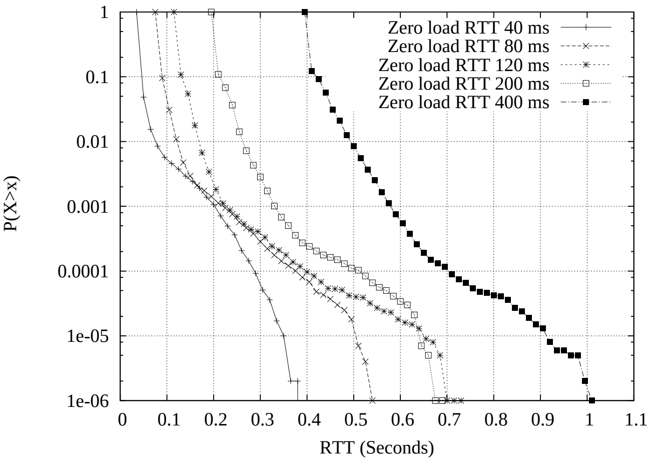


Figure 2. RTT estimation using the “Timestamp” method

40, 80, 120, 200 and 400 ms in the experiments. In order to emulate the behavior of middleboxes, which usually are used in high-performance networks, TCP handshake packets (first 3 packets with SYN, SYN+ACK and ACK flags) will have a lower forwarding delay, exactly a 10% less than the value used for the rest of the connection.

In order to have a scenario of TCP not filling the path $bandwidth \times delay$, TCP window scale option is deactivated.

The experiments consist on clients making HTTP, Hypertext Transfer Protocol, requests to the Web server. The server will send a variable size page whose size follows a uniform distribution between 1 and 3 MBytes. Clients will make requests with inter arrival times following a uniform distribution with mean 8 seconds. These characteristics will provoke that the channel of the server will have an average load about 6 Mbps.

IV. Results

The RTT of a path used by a TCP connection can be defined as the time it takes for a packet to travel from one endpoint to the other plus the time it takes the confirmation packet to return back to the original endpoint. That is a property of the path which could be calculated by just adding the link delays of the path. But an actual TCP connection is affected by the actual RTT of every packet it sends which is not always the pure RTT of the path because of variations due to waiting at queues along the path or response waiting times at the remote endpoint. Thus, the RTT can be seen as a random process. The estimation algorithms are trying to measure this random variable.

In the presented secenario, different RTT estimators have been evaluated, namely the three proposed RTT1, RTT2, RTT3 estimators provide a value for the RTT seen by a given TCP connection. They have been compared to two classical estimators of RTT for TCP connections: *initial RTT* estimator and *TCP timestamp option* estimator. Initial RTT estimator measures the RTT for the connection as the time from first SYN packet of the conception to the confirmation packet of the SYN+ACK packet. This is the time duration of TCP 3way handshake. The TCP timestamp estimator measures RTT of the connection by observing TCP timestamp options that provide accurate instant RTT samples. These samples are always greater than the actual RTT. The TCP timestamp estimator chooses the smallest sample value observed as the RTT for that connection path. The TCP timestamp option estimator will be a very good estimator by definition but its use depends on the captured connections using it.

Apart from that, as every TCP connection in the validation scenario has the same path, the full set of TCP timestamp measure samples could be used as a ground truth of the RTT random process. Figure 2 shows the probability density function of RTT for the different configurations.

To compare the estimators, 2000 connections were captured running the explained scenario. The five estimators RTT1, RTT2, RTT3, initial and TCP timestamp, were computed for every TCP connection seen.

The results, the mean, minimum, maximum and variance values, for all the estimators analyzed are shown in Table I. The probability density functions are shown in Figure 3. The results show that all estimators slightly overestimate the actual RTT and that the TCP timestamp is the most precise as expected. The overestimation is larger when the actual RTT to estimate has low values.

Analyzing the results obtained for the rest of the RTT estimators versus the initial RTT, obtained from the handshake, and the amount of time of the Timestamp, it is shown as the majority of the connections had results around the expected RTT, Figure 3. From the three algorithms, RTT1 is the one which overestimate less since it is the less strict.

Since the scenario emulates a network using middleboxes, the initial delay time which was obtained from the 3way handshake should be slightly lower than the delay for the rest of the connection, (10% lower). However, as the server experienced a moderate load the actual initial RTT was usually on the order of the RTT for the rest of the packets. This effect can be observed at the minimum value obtained for the initial RTT for each experiment, as shown in Table I.

Some connections had high values for all the methods except for the Timestamp estimation. During the connection, sometimes the actual packet RTT was the configured scenario value. In a long connection, at least some window flights experienced an RTT larger than the base one, due to the

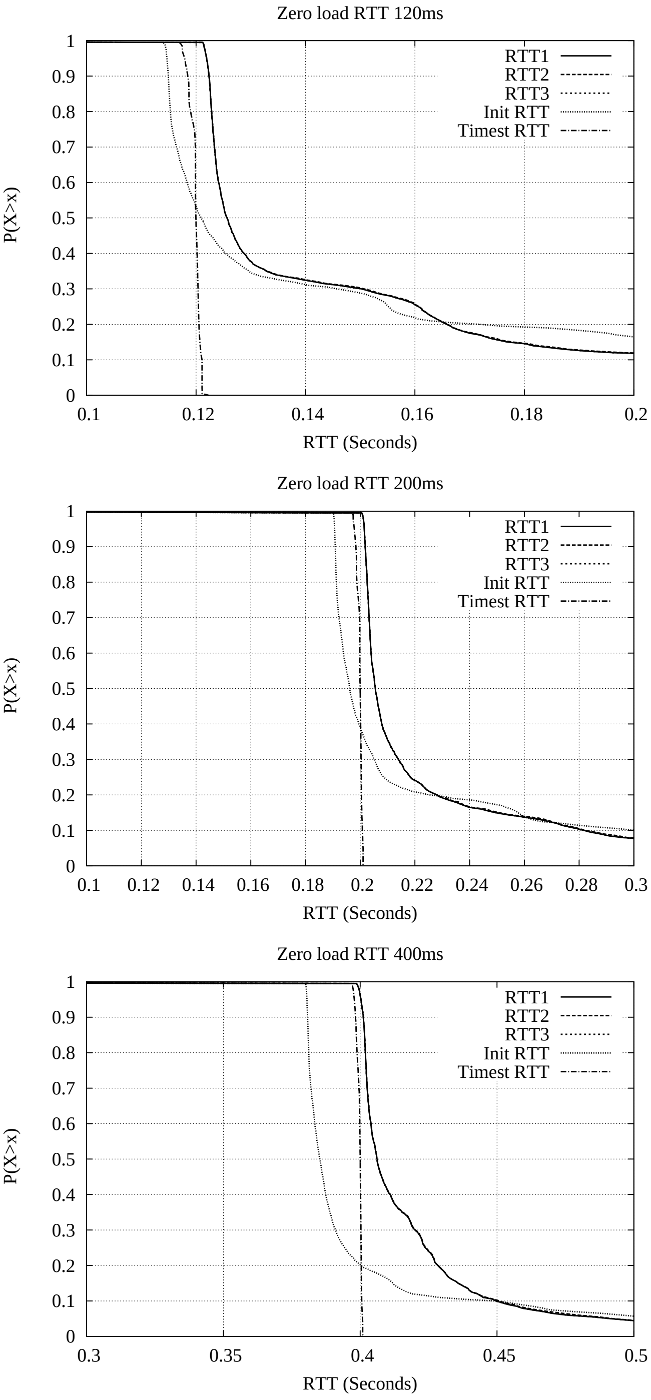


Figure 3. Comparing the RTT estimation using all methods

server being busy with traffic for others requests. The proposed algorithms adapt to the maximum time for the connection and so the values were higher than expected. Figure 4 shows an example for a connection whose RTT values were higher than expected. This connection should have a 400ms estimation for the RTT, according to the theoretical value, however the result of the algorithm for that connection was 700ms. The values obtained were similar to the maximum value for the Timestamp estimation observed for the same connection.

TABLE I. STATISTICS FOR THE CALCULATED RTT

Zero load RTT	Method	Min (s)	Max (s)	Mean (s)	Variance (s^2)
40ms	rtt1	0.054	0.396	0.087	0.003
	rtt2	0.054	0.396	0.087	0.003
	rtt3	0.054	0.396	0.087	0.003
	Initial	0.038	0.417	0.085	0.005
	Timestamp	0.036	0.041	0.040	0.000
80ms	rtt1	0.081	0.569	0.123	0.004
	rtt2	0.081	0.569	0.124	0.004
	rtt3	0.081	0.569	0.123	0.004
	Initial	0.074	0.569	0.126	0.006
	Timestamp	0.076	0.081	0.080	0.000
120ms	rtt1	0.121	0.660	0.151	0.003
	rtt2	0.121	0.660	0.151	0.003
	rtt3	0.121	0.660	0.151	0.003
	Initial	0.113	0.700	0.153	0.005
	Timestamp	0.115	0.122	0.120	0.000
200ms	rtt1	0.200	0.661	0.227	0.003
	rtt2	0.200	0.661	0.228	0.003
	rtt3	0.200	0.661	0.227	0.003
	Initial	0.190	0.712	0.224	0.005
	Timestamp	0.195	0.201	0.200	0.000
400ms	rtt1	0.398	0.945	0.422	0.002
	rtt2	0.398	0.945	0.423	0.002
	rtt3	0.398	0.945	0.422	0.002
	Initial	0.380	0.926	0.404	0.004
	Timestamp	0.396	0.402	0.400	0.000

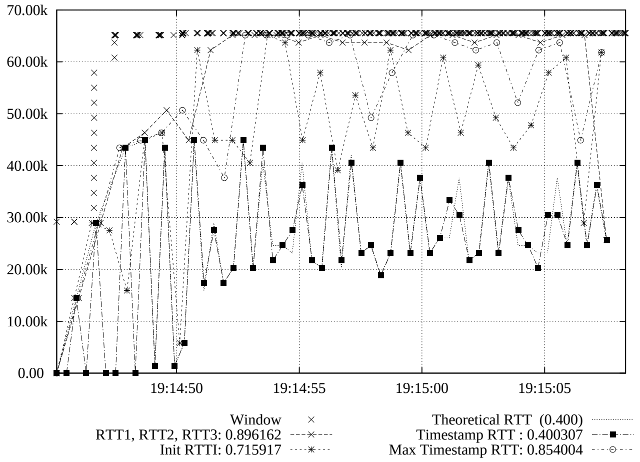


Figure 4. Observed timeseries of bytes for each RTT candidates.

Figure 5 shows the results from the estimators. The mean is slightly lower for the initial RTT compared to the proposed methods. The standard deviation is higher, which indicates a higher variability on the measurement. Besides, it is worth mentioning that for some cases the value obtained from the initial RTT is a subestimation of the real RTT.

Finally, it should be noted that the proposed algorithms check the validity of an RTT candidate by comparing the observed bytes to the advertised window. Thus it needs that the TCP connection is not filling the $bandwidth \times delay$ product for the path. Otherwise the initial candidate RTT is an upper RTT limit, giving no information. The limit is shown in Figure 6 where the minimum RTT for a given bandwidth is plotted. The points $(bandwidth, RTT)$ under the line represent situations where the algorithm does not work. This limit depends on the maximum advertised window allowed by TCP which is 64Kbytes for classical TCP but can be extended if it accepts the window scale option. Figure 6 shows this limit

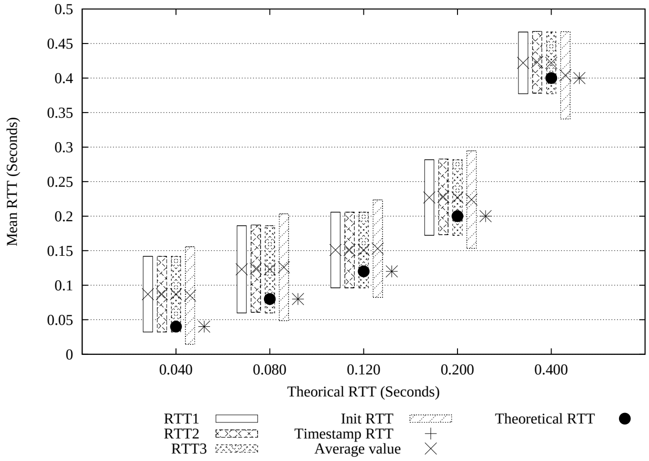


Figure 5. Average RTT and its deviation error obtained for all RTT tests.

for several values of the window scale option, from 64KB to 8MB.

For low bandwidth scenarios, the algorithm can be used to estimate the RTT, provided TCP connections do not use window scaling, or use low values. For higher speed scenarios TCP connections with larger window scale option values can be estimated. For example in a 1Gbps, data center network using window size of 2MB, it is possible to estimate an RTT larger than 16ms. In a longer link with 100Mbps, it is possible to estimate RTTs larger than 40ms provided the window size is 512KB or less. The low bandwidth (10Mbps) scenario used for validation without window scale option allows to estimate RTTs higher than 51.2ms. This can be checked at Table I where the result for RTT1, RTT2 and RTT3 in the 40ms scenario never gives an output lower than 52ms.

Nevertheless even if TCP endpoints agree to use certain window scale value, it does not imply they will advertise the maximum allowed window. Observations at authors' university access link show that even TCP connections usually negotiate window scales allowing up to 8MB advertised windows. However these connections afterwards do not advertise so large windows. Figure 7 shows the survival function of the maximum advertised window used by connections compared to the survival function of the maximum allowed advertised window for the negotiated window scale. Note that around 30% connections negotiate window scales that allow 512KB windows but approximately just 15% actually advertise 512KB. Thus around 85% of the observed link TCP connections will meet the requirements to estimate RTTs larger than 40ms.

V. CONCLUSIONS

The RTT value is a key metric for performance evaluation of a TCP connection. It determines the QoS perceived by the user especially in application level protocols with multiple requests like HTTP. The measurement of RTT is also a requisite for deeper analysis of TCP behavior from passive traffic captures.

However, the calculation of this value is non-trivial in a loaded network as it has to be inferred from packet observed traveling in both directions, taking into account too many parameters such as disorders, retransmission, losses during the capture, etc. RTT can be measured by ICMP probing (Ping) or similar active measurements, but sometimes active injection of traffic is not an option and a passive methodology is preferred.

In this work, a passive methodology is presented to estimate RTT from passive traffic capture. It has been compared

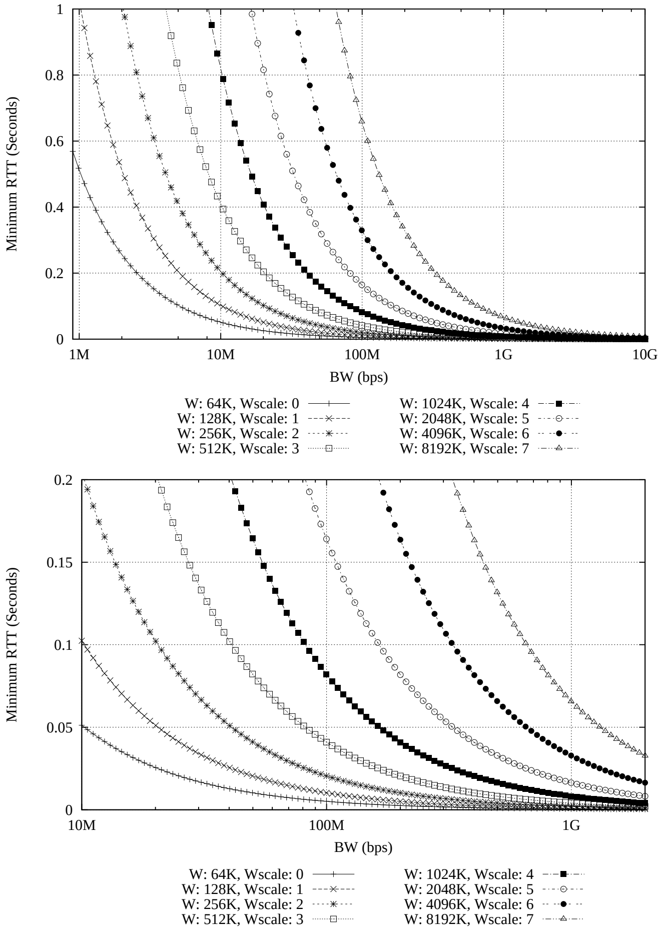


Figure 6. Mean RTT and its deviation error obtained for all RTT tests.

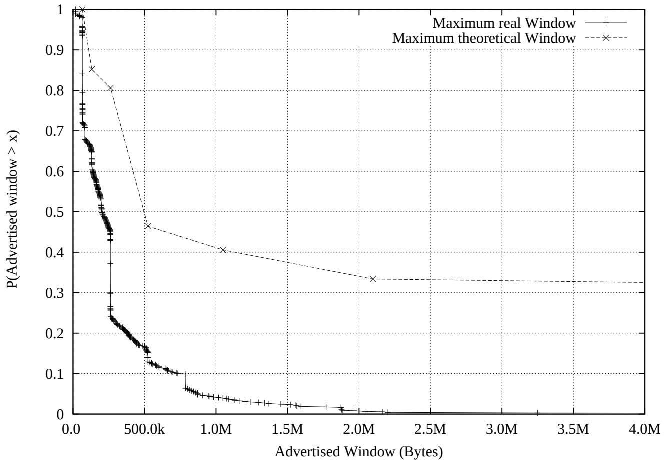


Figure 7. Survival functions of the maximum advertised and allowed windows

to two other classic passive methods: the use of the initial TCP three-way-handshake time, and the observation of the TCP timestamp option defined.

The three estimators are fully passive and can be used on traffic traces.

The proposed algorithm provides an overestimation of the actual RTT value. This is often desired as the values of RTT is used to decide on a timescale above the RTT. The TCP option timestamp estimator has the same property. It never provides a measure lower than the actual RTT but it may give a larger value. On the other hand the initial three-way-handshake RTT may sometimes give smaller RTT samples caused by the presence of middleboxes that answer or establish the connection on behalf of one of the endpoints that is farther away.

It has been shown that in an emulated scenario, the proposed algorithm performs not as accurately as TCP timestamp option method but provides reasonable accuracy. TCP timestamp option method is difficult to improve because it is the observation of an active measurement. The problem with TCP timestamp option method is that the passive estimator requires that the TCP connections are using timestamp option which is not common nowadays.

The proposed algorithm requires that the TCP connection is not filling its $bandwidth \times delay$ product thus, it dependes on the values of RTT and path bandwidth and also on the window scale TCP option but may be used in every connection regardless of its use of timestamp TCP option. Hence it allows the passive estimation of RTT in high bandwidth scenarios (like datacenter networks). In a traffic trace, it provides a RTT estimation for a different set of TCP connections that timestamp option method increasing the number of RTT samples that can be obtained from a captured trace.

ACKNOWLEDGMENT

The authors would want to thank Public University of Navarra for funding through PIF grant.

REFERENCES

[1] L. Zhang, "Why TCP timers don't work well," in Proceedings of the ACM SIGCOMM conference on Communications architectures & protocols, 1986, Stowe, Vermont, United States August 5-7, 1986, 1986, pp. 397–405. [Online]. Available: http://doi.acm.org/10.1145/18172.18216

[2] S. D. Strowes, "Passively measuring tcp round-trip times," Communications of the ACM, vol. 56, no. 10, 2013, pp. 57–64.

[3] "Tcp extensions for high performance. rfc 1323." [Online]. Available: https://tools.ietf.org/html/rfc1323 [accessed: 2015-05-29]

[4] B. Veal, K. Li, and D. Lowenthal, "New methods for passive estimation of tcp round-trip times," in Passive and Active Network Measurement, ser. Lecture Notes in Computer Science, C. Dovrolis, Ed. Springer Berlin Heidelberg, 2005, vol. 3431, pp. 121–134. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-31966-5_10

[5] H. Jiang and C. Dovrolis, "Passive estimation of tcp round-trip times," SIGCOMM Comput. Commun. Rev., vol. 32, no. 3, Jul. 2002, pp. 75–88. [Online]. Available: http://doi.acm.org/10.1145/571697.571725

[6] G. Lu and X. Li, "On the correspondency between tcp acknowledgment packet and data packet," in Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement, ser. IMC '03. New York, NY, USA: ACM, 2003, pp. 259–272. [Online]. Available: http://doi.acm.org/10.1145/948205.948239

[7] S. Jaiswal, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley, "Inferring tcp connection characteristics through passive measurements," in INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, vol. 3, March 2004, pp. 1582–1592 vol.3.

[8] "Netem." [Online]. Available: http://www.linuxfoundation.org/collaborate/workgroups/networking/netem [accessed: 2014-06-01]

Congestion Avoidance TCP Improvements for Video Streaming

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

***Abstract*—Video streaming has become the major source of Internet traffic nowadays. Considering that content delivery network providers have adopted Video over Hypertext Transfer Protocol/Transmission Control Protocol (HTTP/TCP) as the preferred protocol stack for video streaming, understanding TCP performance in transporting video streams has become paramount. In our previous work, we have shown how Slow Start of TCP variants play a definite role in the quality of video experience. In this paper, we research mechanisms within congestion avoidance phase of TCP 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 when playout buffer space is used within TCP congestion avoidance phase.**

***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, as well as rendering stalls due to the lack of timely data. Transport delays and data starvation 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 data transfer 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 [3] [4]. 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 [3], and another using a proportional plus derivative controller [4]. More recently, we have studied TCP performance of most popular TCP variants - Reno [2], Cubic (Linux) [12], Compound (Windows) [13] - as well as our proposed TCP variants: Capacity and Congestion Probing (CCP) [3], and Capacity Congestion Plus Derivative (CCPD) [4], 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 a previous work, we have proposed enhancements on Slow Start phase of TCP to improve video streaming performance [7]. In this paper, we show that it is possible to also alter Congestion Avoidance phase of TCP to improve video streaming over Internet paths with wireless access links. More specifically, we demonstrate that: i) Ensuring minimum throughput above video rendering rate may hurt streaming performance rather than help it; ii) Considering playout buffer size in the congestion avoidance as extra space for TCP packet storage results in consistent performance improvement across various network path scenarios. 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 congestion avoidance schemes 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

Modifications of TCP protocol to enhance video streaming have been recently proposed. Pu et al. [10] 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 congestion avoidance for which congestion window $cwnd$ adjustment is disabled, being 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 [6]. Lu et al. [11] have proposed a receiver based scheme to avoid TCP congestion control in case of lost packets on a wireless link. Our CCP and CCPD variants already differentiate between packet losses due to congestion from wireless link layer losses, their main motivation.

Park et. al. [9] 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.

An analytical framework to the dimensioning of playout buffer has been developed by [14]. The goal is to mitigate buffer underflow as well as packet retransmissions along the path. Our work does not try to dimension the playout buffer, but rather take advantage of its size to improve video streaming performance.

In [8], a relationship between network, application (streaming), and user key performance indicators is studied. They conclude that “rebuffering frequency” impacts the most in user perceived video quality, which is one of our video performance measurers (underflow events).

A distinct aspect of our current work is that we propose improvements on congestion avoidance phase of TCP, and evaluate them on real client and server network stacks that are widely deployed for video streaming, via VLC open source video client, and standard HTTP server.

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. Figure 1 illustrates video streaming components.

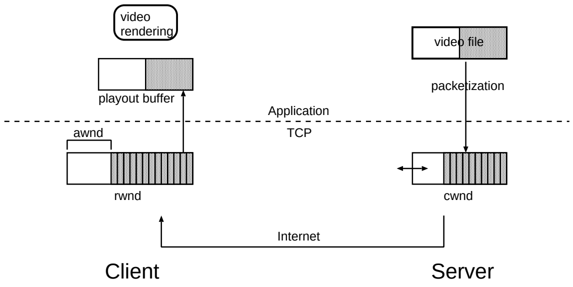


Fig. 1: Video Streaming over TCP

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 congestion avoidance scheme, 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 schemes, 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 [2] and Cubic [12] are examples of additive increase multiplicative decrease (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. Compound [13], CCP [3] and CCPD [4] 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. 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 $cwnd$ size can only increase in this phase. So, there is no flow control of the traffic into the network. This phase ends when $cwnd$ 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 use slow start except Cubic [12].
- **Congestion Avoidance(CA):** This phase is entered when the TCP sender detects a packet loss, or the $cwnd$ size reaches the target upper size $ssthresh$ (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.

Figure 2 illustrates various phases of a TCP session. Our interest is in the congestion avoidance phase of TCP, which dictates how much traffic is allowed into the network during periods of network congestion. A comprehensive tutorial of TCP features can be found in [1].

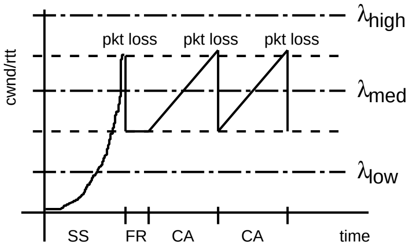


Fig. 2: TCP Congestion Window Dynamics vs Video Playout

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

λ_{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.

λ_{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.

λ_{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.

For most TCP variants widely used today, congestion avoidance phase is sharply different. As we present comparative study of our proposal against Cubic and Compound TCP variants, in what follows we briefly introduce these TCP variants' congestion avoidance phase.

A. Cubic TCP Congestion Avoidance

TCP Cubic is a loss based TCP that has achieved widespread usage as the default TCP of the Linux operating system. During congestion avoidance, its congestion window adjustment scheme is:

$$\begin{aligned} AckRec: \quad cwnd_{k+1} &= C(t-K)^3 + Wmax \\ K &= (Wmax\frac{\beta}{C})^{1/3} \\ PktLoss: \quad cwnd_{k+1} &= \beta cwnd_k \\ Wmax &= cwnd_k \end{aligned} \tag{1}$$

where C is a scaling factor, Wmax is the cwnd value at time of packet loss detection, and t is the elapsed time since the last packet loss detection (cwnd reduction). The rational for these equations is simple. Cubic remembers the cwnd value at time of packet loss detection - Wmax, when a sharp cwnd reduction is enacted, tuned by parameter β. After that, cwnd is increased according to a cubic function, whose speed of increase is dictated by two factors: i) how long it has been since the previous packet loss detection, the longer the faster ramp up; ii) how large the cwnd size was at time of packet loss detection, the smaller the faster ramp up. The shape of Cubic cwnd dynamics is typically distinctive, clearly showing its cubic nature. Notice that upon random loss, Cubic strives to return cwnd to the value it had prior to loss detection quickly, for small cwnd sizes.

Cubic fast release fast recovery of bandwidth makes it one of the most aggressive TCP variants. Being very responsive, it quickly adapts to variations in network available bandwidth. However, because it relies on packet loss detection for $cwnd$ adjustments, random packet losses in wireless links may still impair Cubic's performance.

B. Compound TCP Congestion Avoidance

Compound TCP is the TCP of choice for most deployed Wintel machines. It implements a hybrid loss/delay based congestion avoidance scheme, by adding a delay congestion window dwnd to the congestion window of NewReno [13]. Compound TCP cwnd adjustment is as per 2:

$$\begin{aligned} AckRec: \quad cwnd_{k+1} &= cwnd_k + \frac{1}{cwnd_k + dwnd_k} \\ PktLoss: \quad cwnd_{k+1} &= cwnd_k + \frac{1}{cwnd_k} \end{aligned} \tag{2}$$

where the delay component is computed as:

$$\begin{aligned} AckRec: dwnd_{k+1} &= dwnd_k + \alpha dwnd_k^K - 1, \text{if } diff < \gamma \\ & dwnd_k - \eta diff, \qquad \text{if } diff \geq \gamma \\ PktLoss: dwnd_{k+1} &= dwnd_k(1-\beta) - \frac{cwnd_k}{2} \end{aligned} \tag{3}$$

where α, β, η and K parameters are chosen as a tradeoff between responsiveness, smoothness, and scalability.

Compound TCP dynamics is often dominated by its loss based component. Hence, it presents a slow responsiveness to network available bandwidth variations, which may cause playout buffer underflows.

C. Capacity and Congestion Probing TCP

In this paper, we use CCP as a framework upon which we design congestion avoidance variation schemes. TCP CCP was our first proposal of 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 or decrease. In addition, packet loss does not trigger any special cwnd adjustment. CCP cwnd adjustment scheme is as per 4:

$$cwnd_k = \frac{[Kp(B - x_k) - in\_flight\_segs_k]}{2} \quad 0 \leq Kp \qquad (4)$$

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.

V. TCP Congestion Avoidance Improvements for Video Streaming

The original idea of congestion avoidance was to maintain $cwnd$ at large values without incurring in packet losses, so as to incur in highest throughput possible. However, for video applications, the ideal throughput should not deviate much from the video rendering rate, or else playout buffer underflow or frame discards may happen. For instance, there is no use in aiming at too high throughput, as packets belonging to frames whose playout time is in the future may clog the playout buffer. We introduce a couple of changes in congestion avoidance of our CCP TCP variant:

- **LimitedCongestionAvoidance:** In this scheme, our TCP variant (CCPLCA) in congestion avoidance computes its $cwnd_{ccp}$ as per Eq. 4. In addition, it computes the minimum $cwnd_{vr}$ value for which at current packet rtt experienced results on a throughput matching the video rendering rate ($cwnd_{vr} = VR/rtt$). The exercised $cwnd$ results to be the largest one, or $cwnd = MAX(cwnd_{vr}, cwnd_{ccp})$. The rational is to not allow the regulated throughput to ever go below the video rendering rate.
- **LargeBuffer:** In this scheme, TCP variant (CCPLB) uses the playout buffer length as part of its $cwnd$ computation, as follows:

$$cwnd_k = \frac{[Kp(B - x_k) - in\_flight\_segs_k]}{2} + \frac{POB}{POBRate} \quad 0 \leq Kp \qquad (5)$$

where POB is the playout buffer size, and $POBRate$ represents a percentage of the playout buffer size used in the TCP congestion avoidance phase. The rational is to use the extra space of the playout buffer to increase throughput, reducing buffer underflow events, as well as decrease throughput when playout buffer is close to be full, avoiding frame discards.

VI. Video Streaming Performance of Congestion Avoidance Schemes

Figure 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 congestion avoidance schemes on video streaming performance.

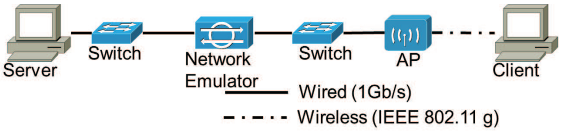


Fig. 3: Video Streaming Emulation Network

TCP variants used are: Cubic, Compound, CCP, CCPLBA, and CCPLB. Performance is evaluated for various round trip time path scenarios, as per Table I.

TABLE I: EXPERIMENTAL NETWORK SETTINGS

Video Size	409Mbytes
Playout time	10.24 secs
Encoding	MPEG-4
Video Codec	H.264/AVC
Audio Codec	MPEG-4 AAC4
Video Playout Buffer Size	448, 897, 1345 pkts
Network Delay (RTT)	3, 100, 200 msecs
TCP variants	Cubic, Compound, CCP, CCPLCA, CCPLB

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.

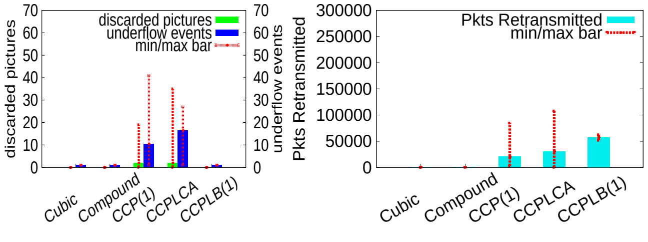
Performance measurers adopted, in order of priority, are:

- **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 experimental results into the following: i)TCP variants performance comparison; ii)CCPLB sensitivity analysis. Each data point in charts represents five trials. Results are reported as average and min/max deviation bars.

A. TCP Variants Performance Comparison

Figure 4 reports on video streaming and TCP performance under short propagation delay of 3msec. In this case, legacy TCP variants Cubic and Compound deliver best video streaming performance with no discarded frames and very small number of playout buffer underflow events. CCP(1), our previous TCP variant, presents significantly more frame discards, as well as buffer underflow events. Even though CCP uses path storage capacity to regulate its input traffic, CCP ignores playout buffer depth. CCPLCA presents worst performance, which shows that simply being liberal in sizing $cwnd$ to large values may end up hurting video streaming performance, rather than helping. One needs to size $cwnd$ to large values only when the playout buffer is able to accommodate the traffic, and quickly use the extra packets to render frames on a timely manner. Finally, our new CCPLB(1) TCP variant (1 means full size of the playout buffer is used) delivers as good a performance as Cubic and Compound legacy TCPs, even though it retransmits more packets than all other TCP variants.

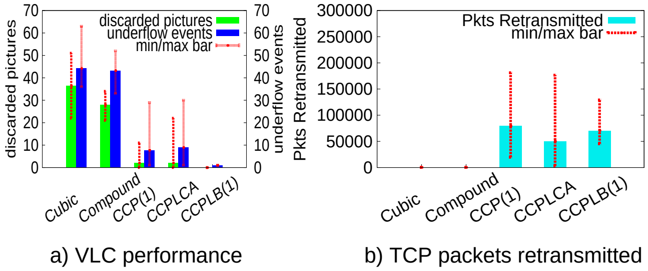


a) VLC performance b) TCP packets retransmitted

Fig. 4: Video Performance vs TCP performance; rtt=3msec

Figure 5 reports on video streaming and TCP performance under a typical propagation delay of 100msec. In this case, legacy TCP variants Cubic and Compound deliver worst video streaming performance among all TCP variants studied. CCP(1), our previous TCP variant, presents significantly less frame discards than the legacy ones, as well as buffer underflow events. Among all variants, CCPLB(1) performs best by maintaining a very low number of playout buffer underflow events, as well as no frame discards. CCPLB(1) is able to keep low number of underflow events and frame discards by taking into account the size of the playout buffer when regulating $cwnd$ window, even though it does not know the instantaneous filling level (number of packets) of the playout buffer. Notice also that the number of retransmitted packets of CCP and CCPLB are roughly the same, even though CCPLB delivers better video performance.

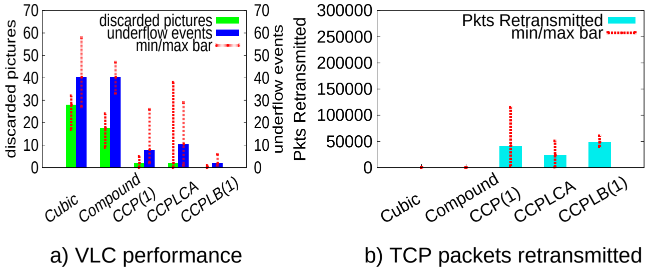
Figure 6 reports on video streaming and TCP performance under a large propagation delay of 200msec. Delays such as that may be experienced in paths with cellular network access links, where additional delays result from wireless access link level retransmissions. In this case, legacy TCP variants Cubic and Compound still deliver worst video streaming performance among all TCP variants. CCP(1) continues to present significantly less frame discards than the legacy ones, as well as buffer underflow events. In addition, CCPLB(1) performs best by maintaining a very low number of playout buffer underflow events, as well as no frame discards, even in the face of a very large round trip delay.



a) VLC performance b) TCP packets retransmitted

Fig. 5: Video Performance vs TCP performance; rtt=100msec

In conclusion, CCPLB is able to consistently deliver best video streaming performance across a wide range of round trip delay paths.



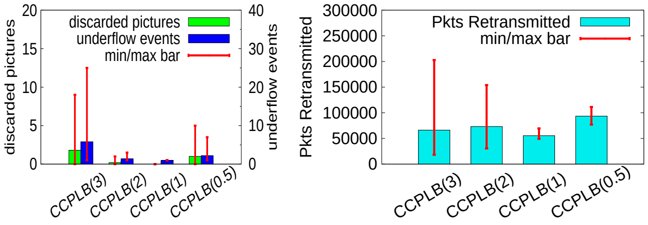
a) VLC performance b) TCP packets retransmitted

Fig. 6: Video Performance vs TCP performance; rtt=200msec

B. Playout Buffer Size Sensitivity Analysis

So far we have presented CCPLB results using the whole playout buffer size. Next we address performance sensitivity to two issues: playout buffer size itself, and the percentage of the playout buffer size used by CCPLB.

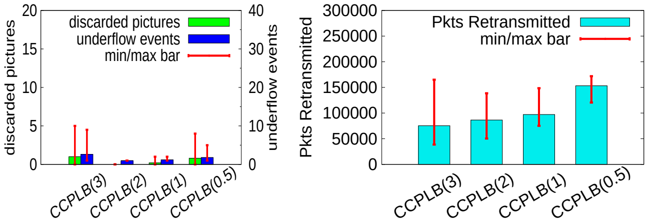
Figure 7 reports on video streaming and CCPLB performance under a typical propagation delay of 100msec and playout buffer size of 448 max size IP packets of 1600 bytes for various amounts of buffering. First notice the small amounts of picture discards, as well as underflow buffer events across all variants. CCPLB(2) uses half the buffer size of the playout buffer in its congestion avoidance $cwnd$ regulation, whereas CCPLB(0.5) uses twice as much buffer as the size of the playout buffer. The later case represents an overbooking of playout buffering, as CCPLB uses more buffering than it is really available at the client. We can see that overbooking hurts performance, whereas underbooking, using less buffering than the total playout buffer size, does not affect video streaming performance significantly. All variants present a reasonable amount of retransmitted packets at the TCP layer.



a) VLC performance b) TCP packets retransmitted

Fig. 7: Video Performance vs TCP performance; POB=448pkts

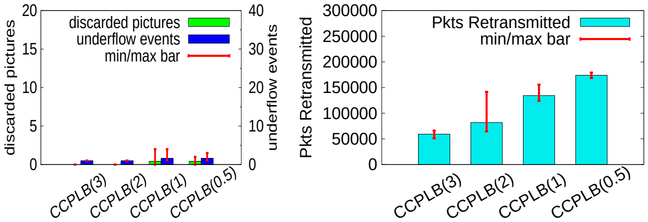
Figure 8 reports on video streaming and CCPLB performance under a typical propagation delay of 100msec and playout buffer size of 897 max size IP packets of 1600 bytes for various amounts of buffering. Comparing CCPLB VLC performance with previous case, there is much less variation in discarded frames as well as underflow events, with half the playout buffer. There is also roughly the same level of packet retransmissions at the TCP level performance from the previous case.



a) VLC performance b) TCP packets retransmitted

Fig. 8: Video Performance vs TCP performance; POB=897pkt

Finally, Figure 9 reports on video streaming and CCPLB performance under a typical propagation delay of 100msec and a large playout buffer size of 1345 max size IP packets of 1600 bytes for various amounts of buffering. Comparing CCPLB VLC performance with previous results, there is no significant improvement in VLC performance. This shows that beyond a certain size, there is no appreciable gain in increasing playout buffer size.



a) VLC performance b) TCP packets retransmitted

Fig. 9: Video Performance vs TCP performance; POB=1345pkt

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. We have simply used parameter values from our previous study of CCP performance of file transfers [5]. Finally, changes to the congestion avoidance phase of CCP can be equally applied to CCPD TCP variant.

VII. Conclusion and Future work

In this paper, we have introduced and evaluated a couple of variations of the congestion avoidance phase of our TCP protocol variant CCP to improve TCP transport performance of video streams. We have characterized CCP performance with these schemes when transporting video streaming applications over wireless network paths via open source experiments. Our experimental results show that taking into account playout buffer size in the regulation of congestion window $cwnd$ results in better video streaming experience, with fewer frame discards as well as less video rendering stalls, across a wide range of path round trip times. As future work, we are currently exploring how playout buffer size may be estimated by the video server. We are also researching how video streaming over multiple paths may affect video rendering experience.

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] 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.

[4] 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.

[5] H. Ishizaki, K. Kumazoe, T. Ikenaga, D. Cavendish, T. Masato, Y. Oie, “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.

[6] 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.

[7] G. Watanabe, K. Kumazoe, D. Cavendish, D. Nobayashi, T. Ikenaga, and Y. Oie, “Slow Start TCP Improvements for Video Streaming Applications,” IARIA Sixth International Conference on Evolving Internet, best paper award, June 2014, pp. 22-27.

[8] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, “Measuring the Quality of Experience of HTTP Video Streaming,” Proceedings of IEEE International Symposium on Integrated Network Management, Dublin, Ireland, May 2011, pp. 485-492.

[9] J-W. Park, R. P. Karrer, and J. Kim,, “TCP-Rome: 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.

[10] 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.

[11] Z. Lu, V. S. Somayazulu, and H. Moustafa, “Context Adaptive Cross-Layer TCP Optimization for Internet Video Streaming,” In Proceedings of IEEE ICC 14, 2014, pp. 1723-1728.

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

[13] 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.

[14] J. Yan, W. Muhlbauer, and B. Plattner, “Analytical Framework for Improving the Quality of Streaming Over TCP,” IEEE Transactions on Multimedia, Vol.14, No.6, December 2012, pp. 1579-1590.

AccessWeb Barometer

A Web Accessibility Evaluation and Analysis Platform

Ramiro Gonçalves [a, b], José Martins [a, b], Frederico Branco [a, b], Jorge Pereira [a], Carlos Peixoto [a], Tânia Rocha [a, b]
[a]University of Trás-os-Montes e Alto Douro
Vila Real, Portugal
[b]INESC TEC and UTAD
Vila Real, Portugal
ramiro@utad.pt, jmartins@utad.pt, fbranco@utad.pt, Jorge.pereira@infosistema.com, carlospeixoto76@hotmail.com, trocha@utad.pt

***Abstract*—The constant evolution of all Web related technologies, and the considerable adoption of these technologies in our society's everyday life, has brought to the discussion table the ability of these Web technologies and Web contents to become accessible to all, including those with some sort of disability. During the past years, a research project has been executed in order to, not only give the Web accessibility topic more visibility within our society, but also to achieve indicators on the levels of accessibility presented by privately held company websites. Considering the growing need to rapidly achieve Web accessibility indicators, whose complexity has significantly increased, the research team inherent to the referred project developed a software platform, entitled "AccessWeb Barometer", that has the ability to perform Web accessibility evaluations to multiple websites in simultaneous. It also has the ability to analyze and publish the results inherent to those evaluations, and to allow its users to create their own analysis and dashboards. In this paper, we present the AccessWeb Barometer software platform architecture, its overall characterization and validation, and also the possibilities of what a platform like this can bring to Web content developers and to organizations worldwide.**

Keywords- Web Accessibility; AccessWeb Barometer; Evaluation Platform; Analysis Service; Diagnostic Service.

I. INTRODUCTION

The topic of Web accessibility has been of major relevance for the global community, and particularly for those who have some sort of impairment or disability. When analyzing the majority of existent websites one can recognize that their compliance levels with current Web accessibility standards is incredibly low [1-3]. The referred topic can be simultaneously seen as an ethical and social problem, but also as an economically relevant issue. By merging these facts with current economic and financial difficulties assumed by almost all organizations, one can pinpoint the importance of a project that focus on, not only identifying websites accessibility issues, but also on providing information on how to solve those same issues.

We present a three layer architecture proposal for a software platform whose goal is to be able to simultaneously evaluate several websites against international Web accessibility, usability and compliance standards, and simultaneously create analytic dashboards that will be made available to all Internet users through a set of collaborative Web platforms.

The present paper is divided into five sections, starting with an introduction section where a very brief approach to the paper's main topic is made. A second section presents the readers with a detailed perspective on the theoretical background inherent to both the relevance of Web accessibility topic and Web accessibility evaluation tools and systems already present in the literature. In third section, a comprehensive description and characterization of the proposed software platform design is made. A fourth section was developed in order to address the validation tests performed in order to ensure reliability to the proposed software solution. The paper finalizes with a fifth section containing some conclusions on the performed work, and on the expected future work.

II. WEB ACCESSIBILITY BACKGROUND CHARACTERIZATION

A. Conceptual Framework

The internet offers a variety of information that, by nature, is constantly changing and evolving, both in size and in complexity, thus becoming an indispensable tool for individuals and organizations in everyday life [4].

Though the Internet is to be used by all, there is a niche of individuals whose physical and/or mental characteristics increases the level of difficulty associated with the referred interaction. Despite their limitations, these individuals should be allowed access to the Web and all its resources in the same manner as a normal user [2]. With this concern in mind, Babu and Sekharaiah [1] argue all Web resources need to incorporate accessibility characteristics that allow disabled users to use them by themselves or by using assistive technologies to do so.

Gonçalves, et al. [5] presented the term "accessibility" as the ability that allows people with some sort of disability or incapability to interact with any product, resource, service or activity in the same manner as an individual without any impairment would. Complementarily, Henry [6] argues that "Web accessibility" is the term used to characterize the ability possessed by Web interfaces that allows them to be perceived, understood, navigable and easy to interact with. Recently, several authors [3, 7, 8] also complemented this initial definition by assuming that it represents Internet usage by

ISBN: 978-1-61208-435-0

everyone, regardless of their physical, perceptive, cultural or social capacities or skills.

According to Gilbertson and Machin [9], there are two parallel approaches one can make to study and work the Web accessibility topic: 1) a more functional approach that focuses on the user's limitations and on the possible solutions (within the available technology) for those limitations; and 2) a more technical approach that focuses on Web technologies and how they can be used, modified or created to diminish or eliminate the obstacles opposing the users to fully benefit from the potential associated with the Web.

B. *Legal and Regulatory Concerns*

In the last two decades, the Web accessibility topic has been on the agenda of several national and international regulatory entities, which allows to highlight the importance of the topic [2].

In recent years, several organizations have been working on the Web accessibility topic. The most prominent one is the World Wide Web Consortium (W3C), mainly due to its Web Accessibility Initiative (WAI) and its Web Content Accessibility Guidelines (now in a second and more updated version) [10]. These guidelines are a set of detailed descriptions to accessibility issues associated with the development of Web applications and content that everyone can use [11]. The current version of the referred guidelines were defined according to several layers of conceptualization, including: principles, general strategies, testable success criteria, a collection of techniques to promote the Web accessibility topic, and a set of complex documentation on all the possible accessibility faults and errors [12].

In parallel with W3C, the International Organization for Standardization (ISO) has been aiming their activities on improving the knowledge inherent to the Web accessibility topic, and to establish a set of standards that should bring the much needed normalization to the area. The most public results of ISOs work have been the ISO TS-16071, ISO 9241-111 and ISO 9241-171 standards that aimed on, not only implementing a set of rules that should be fulfilled, but also helping both the public and organizations to create accessible Web platforms, websites and Web content [13-16].

Despite the existence of several international standards and regulations focused on the Web accessibility topic that were adopted by the majority of the countries, some of them decided to create their own regulations and enforce them at their own will. An example of this creation is the Web accessibility regulation by the United States of America, entitled "Section 508", which highlights the existing right for all data or information be made available by any ICT related systems or any Web platform, to be accessible to all citizens, including those with some sort of disability or incapability [17, 18].

Even though there are several legal and/or regulatory requirements, W3Cs WCAG 2.0 is the most relevant Web accessibility standard. This is the one that most countries and organizations have adopted as the basis for developing accessible Web content.

C. *Recent Perspectives*

According to Burger [19], global tendencies towards increasing Web accessibility levels have significantly improved. Several major software houses and Web consulting agencies are now incorporating accessibility concerns in all their products and contents. The referred author also highlighted that several researchers are also focusing their research activities into, not only developing the technologies in order for them to become more accessible, but also into creating and improving the existing Web development platforms and technologies. This helps developers create accessible Web content and also to promote the Web accessibility topic in both the scientific community and to the general population.

In 2012, Rocha, et al. [20] performed a research project that was aimed at understanding the social and economic reality of individuals that presented some sort of disability or incapability. With this study, these authors were able to conclude that the great majority of the analyzed individuals are unemployed or don't have a factual economic activity, but receive monetary governmental complements and subventions in order to survive. By acknowledging this fact, one can perceive that organizations who do not implement accessible websites are directly neglecting a market share that, due to their impairments, are prone to adopt and use such websites.

Braga, et al. [21] performed a research project in which the authors intended to evaluate the accessibility levels of Bank of Brazil's online banking system. The research was done by using a manual evaluation process that allowed them to better identify the barriers and struggles posed to the referred system users. Assuming that the proposed evaluation methodology was correctly defined, after performing the evaluation activities, the authors were able to acknowledge that some changes were needed and had to be implemented in order for their methodology to be totally usable and reliable. Nevertheless, through the execution of this project, a set of important accessibility faults and issues was identified and transmitted to the bank's IT department in order for them to incorporate the necessary changes.

As stated by Oh and Chen [22], Web accessibility represents an increasingly important variable within the organization's corporate and social responsibility scope. An organization collaborator can perform a decisive part in enforcing both the need to create accessible Web content and presenting an accessible website. With this concern in mind, Santarosa, et al. [23] proposed an accessible e-learning platform that complied with W3C WCAG 2.0, aiming on allowing for universities to offer their students a change, in concerns to the access of information on their courses or classes. In their work, the authors also present strategies to train teachers and educators in order for them to be able to create accessible learning content.

Evaluating websites against Web accessibility standards is not an easy task; the present time surrounds itself with a significant margin for individual or manufacturer interpretation [24]. As reasoned by W3C, when assessing websites accessibility levels one should use a mixed approach and combine both automatic and manual testing in order to guarantee a significant level of reliability [25].

There are several tools to perform Web accessibility assessments in an automated or semi-automated manner, but these tools lack the necessary combination of both a machine perspective and a human comprehension, thus tending to not responding to both the users and the Web content development firms needs [7, 26-28].

III. AccessWeb Barometer – A Web Accessibility Evaluation and Analysis Software Platform

With the AccessWeb Barometer software platform, the research team envisioned to simultaneously create a diagnostic tool that delivered accurate and easy to analyze results, and to raise awareness on the accessibility and usability practices inherent to the design of corporate websites. Execution of website accessibility diagnostics on a large-scale represents a very considerable challenge since the known test instruments are manual or semi-automated, and require the allocation of an unsustainable amount of human resources in order to ensure an acceptable execution time [29].

From the experience collected from previous research and development projects, the research team was able to perceive that each website evaluation takes an average of 6 hours to be evaluated by the software tools. After that, another 40 minutes of specialized work, performed by an expert, in order for the evaluation results to be analyzed. With this in mind, the proposed system allows the execution of a great number of simultaneous evaluations in a smaller period of time. Above all, it increases the degree of confidence in the results, by eliminating the error inherent to human intervention in the analysis of the evaluation data.

Besides the diagnostic and analytics layers of the proposed software platform, another very important part is the Website component because it represents the platform's public interface where users can become more aware of the Web accessibility topic, and interact with the various outputs and results from all the performed Web accessibility evaluations. With this component users can, in a collaborative manner, acquire several new information and resources on the Web accessibility topic. Users can also perform synchronous and asynchronous discussions with other users and with the platform administrators or moderators.

A. Proposed System

The proposed architecture for the evaluation and analysis platform is composed of three different layers (Figure 1), with two of them representing the back end (responsible for the diagnosis and analysis - Diagnostic Layer; Analytics Layer) and the other one representing the Website Layer and serving as an accessible front end.

A three tier system was defined and implemented in order to address a fault very much present in the everyday life of those who are responsible for developing Web content and platforms, and to those who are facing the need to have accessible websites in order to benefit from its content. By allowing for a full automatic mechanism that only needs a list of websites to start evaluating them and to publish their results in a modern and dynamic manner that simply allows users to create their own results analysis and achieve more personal acknowledgements.

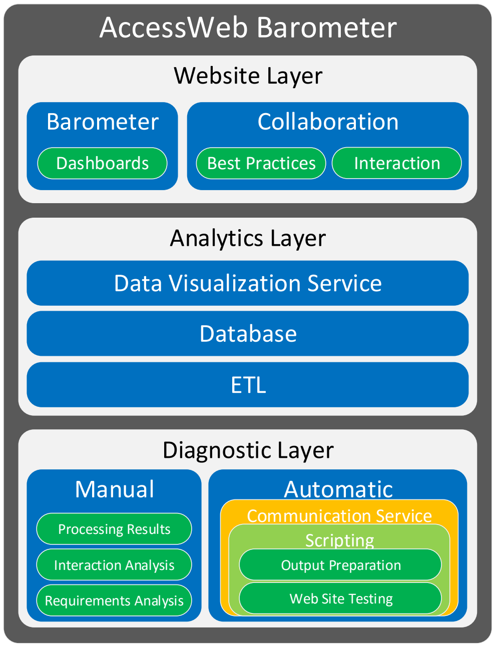


Figure 1. AccessWeb Barometer Software Platform Architecture.

At each layer, there are a set of well-defined tasks that need to be performed and that are responsible for delivering input to the components of the upper layer.

In the following sections, we describe in detail the intrinsic function of each component that integrates the AccessWeb Barometer platform architecture.

B. Diagnostic Layer

In the proposed architecture, the diagnostic layer represents all the components responsible for the accessibility evaluations that is to be performed during the execution of the project inherent to the AccessWeb platform. All Web accessibility evaluations will be supported by W3C WCAG 2.0 and will follow the indications from W3C and use both automatic and manual evaluation mechanisms and techniques [12]. The automated tools are usually fast, but are not able to identify all existent accessibility, usability and compliance issues. Thus, there is a need to complement these automatic assessments with manual reviews. This helps to ensure issues such as language clarity and navigation ease.

The proposed platform incorporates both the use of automated evaluation tools and manual reviews with real users in real environments. This aims on achieving a unified model for analyzing and reaching conclusions on the real limitations that a given website might pose to its users.

1) Manual Evaluation

In the first architectural layer, the Diagnostic Layer, we will focus on the assessment of websites by inspecting their compliance with international guidelines, which are presented to a specialist, and an evaluator. They verify if the system complies with each guideline and registers all failures observed. During the manual evaluation stages the research team will include in the evaluation activities real users and evaluators. The main objective of presenting new and more hands-on results can complement the ones achieved during the parallel automatic evaluation procedure. Therefore, the evaluators will use a manual direct review approach to proceed with the inspection of compliance with Web accessibility and usability guidelines. On the other hand, real users will also be included in the evaluation procedures and through direct interaction, will explore and assess the tested websites interfaces [30-32].

In this context, we will use the barrier's walkthrough method and set-up the following stages of assessment [33]: 1) Identification of scenarios involving two types of users (with visual and motor disability); 2) Definition of accessibility evaluation objectives; 3) Execution of scenarios identified; 4) Analysis of the results; and 5) Presentation of a list of problems with severity level for each of the problems identified by the evaluator.

The assessment of accessibility is not complete without an additional usability evaluation; therefore, we will follow the criteria for measuring usability established by the ISO 9241 standard: 1) Analysis of the characteristics required of the product in a specific context of use; 2) Analysis of the interaction process between the user and the product / system / design; and 3) Analysis of efficiency (agility in enabling work), effectiveness (guarantee that the planned results are obtained) and satisfaction, resulting from the use of the product [34].

Within this scope, in order to achieve the above criteria, we will apply the following usability evaluation techniques: usability testing, cognitive walkthrough, questionnaires and interviews.

2) *Automatic Evaluation*

There is a wide variety of software and online services that help determine if a given website complies with the existent Web accessibility and guidelines, and also with other technology standards. The AccessWeb Barometer platform was not envisioned to be just another assessment platform, but instead, it aims on providing a public barometer that reveals an extensive set of indicators, in a graphical manner, that encourage discussion on the degree of preparation presented by websites, and on the possible interest to society of having accessible and usable websites and Web content.

Despite the existence of several Web accessibility and usability evaluation software tools, to our knowledge there are no solutions for performing multiple and simultaneous websites evaluations. In the proposed platform, the automatic diagnostic component consists of multiple virtual machines, mounted according to the size of the pool of websites that are going to be evaluated, giving the platform a very interesting scalability level. The limits inherent to this approach lie on the physical resources presented by the virtualization servers and on the available Internet access bandwidth.

The Communication Service subcomponent running on each virtual machine has the role of orchestrating and commanding all of the evaluation process of a given website. Its first task is to validate the existence of records on the queue containing the websites to be evaluated. This action is done through proper database queries, which return specific websites attributes (such as name and url), that are needed to perform the referred evaluation and make sure that only those who haven't been evaluated yet are queued. Each website evaluation is launched at the same pace that the virtual machine becomes available. This ensures that a given website is only evaluated once and by a single machine. At the same time the process starts, an update to the website database record is made, in order to "mark it" as already in evaluation. By being aligned with these procedures, the Communication Service subcomponent passes the website url parameter to the scripting application, named "AutoIt 2015", which is responsible for coordinating the execution of the software that will be used to conduct the evaluations.

AutoIt scripting tool runs on each virtual machine, serving to automate the graphical interface of Windows operating system, (i.e., assumes the user's role and performs all the steps that need to be performed for the site to be properly analyzed and evaluated). The proposed website evaluation software platform also incorporates Power Mapper's "SortSite V5.0" whose aim is to perform website evaluations against international Web related standards, such as Section 508, WCAG 2.0 and usability.gov guidelines.

After SortSite finishes the evaluation of a given website, the scripting tool stores the generated reports, passing the workflow again for the Communication Service subcomponent, which will move the evaluation reports to a shared folder ("dump" folder) on the platform server machine (virtual machine responsible for Extract, Transform and Load (ETL) and Data Analysis/Visualization). This process ends with an update to the database, thus ensuring that the website record will be marked as already evaluated and starting a new website evaluation cycle.

The data treatment inherent to the generated reports is performed by the Analytics Layer components described in the following section.

C. *Analytics Layer*

The intermediate layer, entitled Analytics Layer, of the proposed software platform architecture, runs on the server side. It is responsible for the analysis and process of data collected by the lower layer, and aims to prepare it to be presented in the next layer. In practical terms, the components inherent to this intermediate layer is responsible for the Information and Knowledge needed to feed the dashboards that are going to be displayed to the public through the Website Layer.

The first task of this layer workflow is performed by the ETL component, which is of vital importance since it involves moving data from their original sources into the BI system. The ETL component is used to construct and populate the central data repository of the BI architecture, but it is also for identifying relevant data sources in order to build a stable data model (which uniforms, through metadata, all kinds of data), and organize data according to business policies and data storage [35]. In the proposed architecture, the ETL extracts the evaluation report files, stored in the dump folder, and treats and stores the inherent data into new database records. When this process is finished, the component will move the reports files to a “log folder”, ensuring a copy of the evaluation results and a possible future data recovery.

Given the need to store and access data by almost all of the proposed architecture components, a database component was incorporated in order to serve as a central data repository structure, according to a traditional transactional approach.

For the set of data analysis related tasks, some Self-Service Business Intelligence (SSBI) techniques and technologies were used. The main goal of the SSBI is to assist managers in making decisions based on highly complex data analysis and involving less IT know-how’s. This allows for the common user to add new perspectives to the predefined analysis and produce their own queries and reports. This increase of autonomy allows productivity gains for both regular users and IT departments that don’t need to allocate their elements to provide analytical technical support [36].

Having detected the need for a solution that allowed for the representation of complex data in a graphic form, the research team decided to incorporate a Data Visualization (DV) approach, which by definition allows for the visual representation of data and enhances the value of the available information, allowing for an easy identification of trends, exceptions and deviations that are normally hidden in massive amounts of data stored in data sources [37, 38]. This feature of the proposed platform is highly critical because it uses creativity, design concepts, colors, shapes and sizes, to create visual contents that represent the knowledge inherent to a large amount of data [39]. To support this activity, several DV techniques and technologies were used, such as analytical models and statistical functions, whose results are presented visually through interactive dashboards composed of tables, charts, graphs, diagrams, histograms and maps [40].

In the analytics layer, in order to perform the data analysis and visualization, a decision was made to use the Microsoft’s BI stack (“Microsoft Power BI”), which responds to all requirements specified for this software platform and for the graphical representation of data.

D. *Website Layer*

Website Layer corresponds to the front end of the platform, (i.e., serves to interact with all the users). Although the architecture in this layer is composed by two components, the “Barometer” and the “Collaboration” component, a special attention is given to the Barometer component as it is the one that is directly related to the evaluation system results. As a consequence, the Collaboration component enjoys a certain degree of independence by enabling users’ access to additional resources, such as, documentation on best practices in the areas of accessibility, discussion forums and blogs, which have no direct relation with the evaluation system.

Barometer component serves to share all the knowledge extracted from the multiple evaluations carried out to different organization websites through visually rich dashboards; for example, providing a varied combination of graphics, manipulated by a wide range of filters, according to the users' preferences. These indicators range from the analysis of various sectors of activity, analysis of the most common mistakes, and the geographical distribution faulty websites.

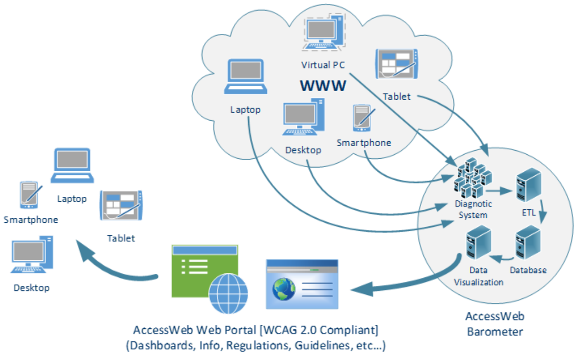


Figure 2. AccessWeb Barometer software platform workflow.

As shown in Figure 2, AccessWeb Barometer software platform acknowledges and also incorporates the prominent role of studying and analyzing the accessibility, usability and compatibility of websites when accessed through the various types of existing devices. All the front-office related components are defined to not only have an attractive and updated design, but also to be responsive (adaptable to both desktop and mobile environments), be compliant with WCAG 2.0 guidelines and compliant with international usability guidelines.

IV. Proposed Platform Validation

Given the complexity associated with the proposed solution, the research team decided that an initial validation stage was needed in order to ensure that, not only all the platform outputs were adequate and correct, but also to acknowledge that an increase of efficiency and performance of the Web accessibility evaluation process was verified.

In order to perform the referred initial validation, and following previous works [5], 1000 Portuguese privately held companies with the biggest business volume were chosen to be used as the evaluation target group. In Figure 3, it is possible to perceive that from the 1000 initial companies, only 862 were evaluated, mainly because the remaining were without a website or had one that was in maintenance or was incompatible with PowerMapper Sortsite tool.

After performing the analysis of the target group and achieving the list of the 862 companies whose websites were to be evaluated, the research team registered that same set of websites in the Analytics Layer “Database” component and started the websites evaluation procedure against WCAG 2.0.

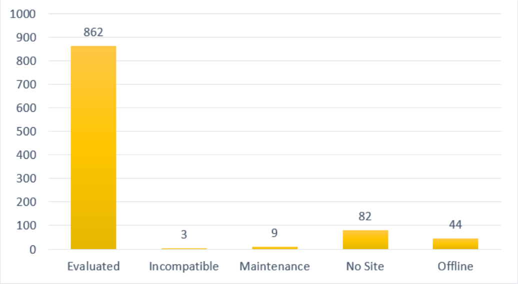


Figure 3. AccessWeb platform initial validation target group analysis.

When all evaluation results were registered in the Analytics Layer by the Diagnostic Layer we were able to achieve all the visual dashboards needed to acknowledge the accessibility levels and compliance presented by the evaluated websites.

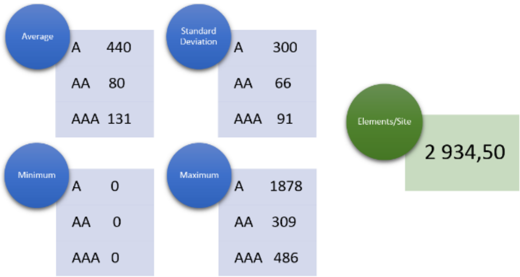


Figure 4. Statistical analysis of the Web accessibility evaluation results.

From the analysis of Figure 4, one can perceive that, despite the average number of evaluated elements from each website is significant ($\approx$ 2900), the average number of detected accessibility errors is still very considerable of what might represent, in line with previous studies [2, 5, 41]. Those levels of compliance with WCAG 2.0 are still very low and those with some sort of disability cannot access the majority of the target group websites without encountering several difficulties or impossible to transpose barriers.

By examining the achieved results, one can acknowledge that the proposed platform is capable of delivering valid and accurate results that allow for a simple and direct understanding on the Web accessibility status of a given website or sets of websites.

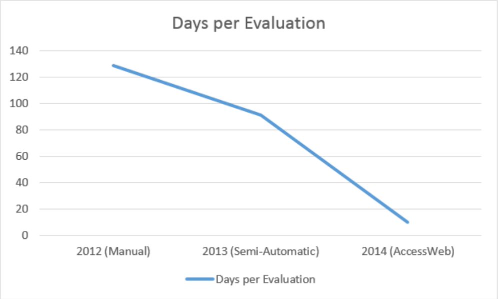


Figure 5. Comparison between the duration of previous Web accessibility evaluations and the one perfomed with AccessWeb platform.

Another critical issue for the research team was the performance presented by the proposed platform. In Figure 5, one can observe a direct comparison between the performance from previous evaluations that the research team performed to the same target group, and the performance presented by this new Web accessibility evaluation performance. From this observation, one can easily highlight the significant improvement of the time necessary for undergoing an evaluation to a set of 1000 companies.

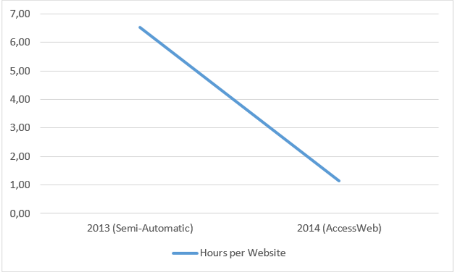


Figure 6. Comparison between the average number of hours necessary for evaluating a website against WCAG 2.0 when using a semi-automatic approach and when using the AccesWeb platform.

By analyzing Figure 6, it is possible to recognize that the use of the AccessWeb platform brings a very interesting improvement to the overall Web accessibility evaluation. Given that, it can reduce in a considerable manner the number of hours necessary to fully evaluate a website, to validate and store the achieved results, and to reach visual dashboards that allow for a direct visualization of the referred website accessibility status.

V. CONCLUSION

An accessible website should allow all users, regardless of their physical or mental situation or impairments, to understand, navigate and interact with the published content. When analyzing the current perception on the Web accessibility concept, one can easily perceives that it is no longer just a technical issue, but also an ethical and social issue, a market (economic) issue, and a SEO issue.

According to W3C one of the most common reasons to the lower levels of accessibility presented by websites is the lack of knowledge which organization managers, Web software developers and Web content creators have; on topics such as Web accessibility standards, assistive technologies and development tools. Drawing on this assumption, the research team inherent to the present project projected a software platform, entitled “AccessWeb Barometer”, for performing multiple accessibility evaluations to sets of websites (mainly belonging to private organizations), giving public access to the results of those evaluations and with this, increasing the global awareness on the Web accessibility topic, and on the importance that it has on the lives of those with some sort of disability or incapability.

With this paper we propose an architecture proposal for the referred software platform that is composed by three main layers, a diagnostic layer (constituted by several components directly responsible for the accessibility evaluations to the chosen websites), an analytics layer whose components have a direct intervention in the extraction of the data inherent to the evaluation processes, in the storing of that same data and in the creation of sets of analyzed and treated information that will be serving as the basis for the public dashboards that are to be available to users through the website layer. This third layer will not only be constituted by the graphic elements that will show the evaluations results to the users, but also by a set of collaborative tools and technologies that should be used to increase the public awareness on the Web accessibility topic.

Currently, all architecture components have been developed and an initial Web accessibility evaluation to a set of 1000 Portuguese company websites was performed, allowing not only to validate that the proposed platform is accurately evaluating the chosen websites, but also to verify that output results are valid and in line with other similar Web accessibility scientific works.

From the referred platform validation, the research team could also identify that the proposed platform ensures a very significant improvement in the overall Web accessibility evaluations field, not only by decreasing the amount of time necessary to perform bulk Web accessibility evaluations, but also by reducing the average time necessary to evaluate a single website.

By incorporating all the considerations achieved from the actions mentioned above, the research team is already planning a future Web accessibility evaluation that focus its attention on websites belonging to both large European companies and SMEs.

ACKNOWLEDGMENT

This paper was supported by the project AcessWeb – Barómetro da Acessibilidade Web em Portugal, financed by the Portuguese Science and Technology Foundation, with an eligible investment of 50.000,00€.

REFERENCES

[1] J. Babu and C. Sekharaiah, "A Panorama of Web Accessibility," *International Journal of Computer Science and Mobile Computing,* vol. 3, pp. 311-317, 2014.

[2] R. Gonçalves, J. Martins, J. Pereira, M. Oliveira, and J. Ferreira, "Enterprise Web Accessibility Levels Amongst the Forbes 250: Where Art Thou O Virtuous Leader?," *Journal of Business Ethics,* vol. 113, pp. 363-375, 2013/03/01 2013.

[3] E. Capra, S. Ferreira, D. Silveira, and A. Ferreira, "Evaluation of Web Accessibility: An Approach Related to Functional Illiteracy," *Procedia Computer Science,* vol. 14, pp. 36-46, 2012.

[4] N. Fernandes, D. Costa, C. Duarte, and L. Carriço, "Evaluating the Accessibility of Web Applications," *Procedia Computer Science,* vol. 14, pp. 28-35, 2012.

[5] R. Gonçalves, J. Martins, J. Pereira, M. Oliveira, and J. Ferreira, "Accessibility levels of Portuguese enterprise websites: equal opportunities for all?," *Behaviour & Information Technology,* vol. 31, pp. 659-677, 2012.

[6] S. Henry, "Understanding web accessibility," in *Web Accessibility*, ed: Springer, 2006, pp. 1-51.

[7] M. Sánchez-Gordón and L. Moreno, "Toward an Integration of Web Accessibility into Testing Processes," *Procedia Computer Science,* vol. 27, pp. 281-291, // 2014.

[8] W3C. (1997, July). *World Wide Web Consortium Launches International Program Office for Web Accessibility Initiative*. Available: http://www.w3.org/Press/IPO-announce

[9] T. Gilbertson and C. Machin, "Guidelines, icons and marketable skills: an accessibility evaluation of 100 web development company homepages," in *Proceedings of the international cross-disciplinary conference on web accessibility*, 2012, p. 17.

[10] S. Henry, "Understanding web accessibility," in *Constructing accessible web sites*, ed: Springer, 2002, pp. 6-31.

[11] Y. Rogers, H. Sharp, and J. Preece, *Interaction design: beyond human-computer interaction*: John Wiley & Sons, 2011.

[12] W3C. (2008, July). *Web Content Accessibility Guidelines (WCAG) 2.0*. Available: http://www.w3.org/TR/WCAG20/

[13] ISO, "9241-110: 2006. Ergonomics of human system interaction-Part 110: Dialogue principles," *International Organization for Standardization (ISO). Switzerland,* 2006.

[14] ISO, "TS 16071: 2003: Ergonomics of human-system interaction–Guidance on accessibility for human-computer interfaces," *International Standards Organisation, Geneva, Switzerland,* 2003.

[15] ISO, "9241-171 (2008) Ergonomics of humansystem interaction--Part 171: Guidance on software accessibility," ed: ISO, 2008.

[16] D. Rømen and D. Svanæs, "Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users," *Universal Access in the Information Society,* vol. 11, pp. 375-385, 2012.

[17] P. Jaeger and M. Matteson, "e-government and technology acceptance: The case of the implementation of section 508 guidelines for websites," *Electronic Journal of e-Government,* vol. 7, pp. 87-98, 2009.

[18] *Section 508*, U. S. Government 2015, 1998.

[19] D. Burger, "Putting e-Accessibility at the Core of Information Systems " *Business Case White Paper Series,* p. 32, 2013.

[20] T. Rocha, M. Bessa, R. Gonçalves, E. Peres, and L. Magalhães, "Web Accessibility and Digital Businesses: The Potential Economic Value of Portuguese People with Disability," *Procedia Computer Science,* vol. 14, pp. 56-64, 2012.

[21] H. Braga, L. Pereira, S. Ferreira, and D. Silveira, "Applying the Barrier Walkthrough Method: Going Beyond the Automatic Evaluation of Accessibility," *Procedia Computer Science,* vol. 27, pp. 471-480, 2014.

[22] L. Oh and J. Chen, "Determinants of employees' intention to exert pressure on firms to engage in web accessibility," *Behaviour & Information Technology,* vol. 34, pp. 108-118, 2015/02/01 2014.

[23] L. Santarosa, D. Conforto, and B. Neves, "Teacher Education and Accessibility on E-Learning System: Putting the W3C Guidelines into Practice," *Teacher Education,* vol. 4, 2015.

[24] V. Centeno, C. Kloos, J. Fisteus, and L. Álvarez, "Web Accessibility Evaluation Tools: A Survey and Some Improvements," *Electronic Notes in Theoretical Computer Science,* vol. 157, pp. 87-100, 5/22/ 2006.

[25] M. Snaprud, K. Rasta, K. Andreasson, and A. Nietzio, "Benefits and Challenges of Combining Automated and User Testing to Enhance e-Accessibility – The European Internet Inclusion Initiative," in *Computers Helping People with Special Needs*. vol. 8547, K. Miesenberger, D. Fels, D. Archambault, P. Peňáz, and W. Zagler, Eds., ed: Springer International Publishing, 2014, pp. 137-140.

[26] A. Iglesias, L. Moreno, P. Martínez, and R. Calvo, "Evaluating the accessibility of three open-source learning content management systems: A comparative study," *Computer Applications in Engineering Education,* vol. 22, pp. 320-328, 2014.

[27] R. Bernard, C. Sabariego, D. Baldwin, S. Abou-Zahra, and A. Cieza, "BETTER-Project: Web Accessibility for Persons with Mental Disorders," in *Human-Computer Interaction: Users and*

Contexts. vol. 9171, M. Kurosu, Ed., ed: Springer International Publishing, 2015, pp. 25-34.

[28] F. Kamoun and M. Almourad, "Accessibility as an integral factor in e-government web site evaluation: The case of Dubai e-government," *Information Technology & People,* vol. 27, pp. 208-228, 2014.

[29] M. Gordon, "Web accessibility evaluation with the crowd: using glance to rapidly code user testing video," in *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, 2014, pp. 339-340.

[30] D. Kreps, "How the web continues to fail people with disabilities," presented at the ALT-C, Leeds, UK., 2008.

[31] L. Law and E. Hvannberg, "Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform," in *Proceedings of the second Nordic conference on Human-computer interaction*, 2002, pp. 71-80.

[32] A. Lepistö and S. Ovaska, "Usability evaluation involving participants with cognitive disabilities," in *Proceedings of the third Nordic conference on Human-computer interaction*, 2004, pp. 305-308.

[33] T. Rocha, "Metáfora de Interação para o Acesso à Informação Digital de uma Forma Autónoma por Pessoas com Deficiência Intelectual," PhD Thesis, Universidade de Trás-os-Montes e Alto Douro, Vila Real, 2014.

[34] ISO, "9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)," *The international organization for standardization,* vol. Parts 1–17, 1998.

[35] S. Bergamaschi, F. Guerra, M. Orsini, C. Sartori, and M. Vincini, "A semantic approach to ETL technologies," *Data & Knowledge Engineering,* vol. 70, pp. 717-731, 2011.

[36] A. Abelló, J. Darmont, L. Etcheverry, M. Golfarelli, J. N. Mazón López, F. Naumann*, et al.*, "Fusion cubes: Towards self-service business intelligence," 2013.

[37] J. Heer, M. Bostock, and V. Ogievetsky, "A tour through the visualization zoo," *Commun. ACM,* vol. 53, pp. 59-67, 2010.

[38] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis," *Queue,* vol. 10, p. 30, 2012.

[39] H. Grierson, J. Corney, and G. Hatcher, "Using visual representations for the searching and browsing of large, complex, multimedia data sets," *International Journal of Information Management,* vol. 35, pp. 244-252, 2015.

[40] D. J. Janvrin, R. L. Raschke, and W. N. Dilla, "Making sense of complex data using interactive data visualization," *Journal of Accounting Education,* vol. 32, pp. 31-48, 2014.

[41] S. Henry, S. Abou-Zahra, and J. Brewer, "The role of accessibility in a universal web," in *Proceedings of the 11th Web for All Conference*, 2014, p. 17.

Session Management of a Variable Video Rate Streaming Session over Multi-Channel Networks

Rachel Behar
Faculty of Computer Science
Jerusalem College of Technology
Jerusalem, Israel
Email: rharris@jct.ac.il

Yael Samet
Intel
Jerusalem, Israel
Email: ygalinsk@jct.ac.il

Ronit Nossenson
Service Performance,
Akamai Technologies
Cambridge, MA, USA
Email: rnossens@akamai.com

***Abstract*—In this paper, we propose and evaluate two algorithms for session management of a variable bit rate video session over a multi-channel network. The session manager decides how many channels should be active in the next time interval on the basis of required video bit rate, session measurement reports and other considerations. The algorithms performance is evaluated against a fixed selection of the number of active channels. Simulation results reveal that it is possible to control the session costs in terms of the number of active channels while keeping the quality of the received video stream on the top Mean Opinion Score (MOS) level. System performance significantly improved in the feedback-based managed session, as compared to the simple-managed session and to the fixed selection of the number of active channels sessions.**

Keywords- multi-channel video transmission; session management; multimedia networks; video QoS.

I. INTRODUCTION

A video streaming application encodes, packetizes and transmits video frames in real-time. In other words, every streaming video frame needs to meet a play-out deadline. Currently, most networks support real-time services only in a best-effort manner. Therefore, video streaming services have to include special measures to be resilient to packet loss and late arrival. Over the last decade, streaming over multiple channels (also called multi-path, or networks) has been suggested to improve the video quality over the Internet [5][6][9]-[14], in peer-to-peer networks [15][19]-[21] and wireless ad-hoc networks [7][8]. Multi-channel video transmission is often coupled with adaptive/scalable layered-video encoding, e.g., H.264, Scalable Video Coding (SVC), to overcome channel rate variation and heterogeneous video client capabilities. Using multiple channels in layered-video transmission has also led to new challenges, such as video packet scheduling and new multi-channel encoding schemes [4][8][16]-[18].

In this study, we explore the contribution of the session management module in a multi-channel variable bit rate video session. Assuming that *M* channels can be activated in a particular video session, the session manager decides on the number of active channels A ≤ *M*. The decision is based on the required video bit rate, and on the session measurement feedback reports regarding the channel conditions from the beginning of the session up to the last time interval. Additional considerations affecting this decision include the required video Quality of Service (QoS) parameters, information regarding the Quality of Experience (QoE) parameters, etc. The number of active channels corresponds to target performance indicators, such as target error rate in the next time interval. We focused on the minimal number of active channels that satisfy the performance requirements in order to reduce the overall system overhead. The session management algorithms described in this study have the following properties: (i) Simple decision function; (ii) Low computation afford; (iii) Small state and storage requirements; and (iv) Little channel feedback information (used only by the second algorithm).

Recently, an algorithm for session management of multi-channel constant rate video streaming session over wireless networks was suggested in [3]. We propose and evaluate two enhanced algorithms that support a variable bit rate video session. The evaluation of the management algorithms is based on simulation environment. The simulation results show that it is possible to control the session costs in terms of the number of active channels, while keeping the quality of the received video stream in the top mean Opinion Score (MOS) level. For example, comparing with static selection of three channels, the simple management algorithm achieved a 13.67% percent cost reduction with 2.59 active channels on average, and average Peak Signal-to-Noise Ratio (PSNR) of 37.32 comparing to average PSNR of 38.14 (reduction of only 2.15%). The feedback-based management algorithm achieved an 8.67% percent cost reduction with 2.74 active channels on average. Furthermore, the feedback-managed algorithm delivered 89.28% bytes on-time compared to only 84.66% bytes that were delivered on-time using a static selection of three channels. This leads to an average PSNR of 38.51, that is, a 0.97% percent improvement in the average PSNR compared to the static selection of three channels. Using a more sophisticated decoder could further improve the PSNR.

This paper is organized as follows: In the next section, the problem statement and rationale for session management are discussed. The simple management and feedback-based management algorithms for a variable bit rate multi-channel video streaming session are described in Section III. In Section IV, we present our simulation environment. In Section V, the simulation results are reported. Conclusions and further research directions are discussed in the last section.

ISBN: 978-1-61208-435-0

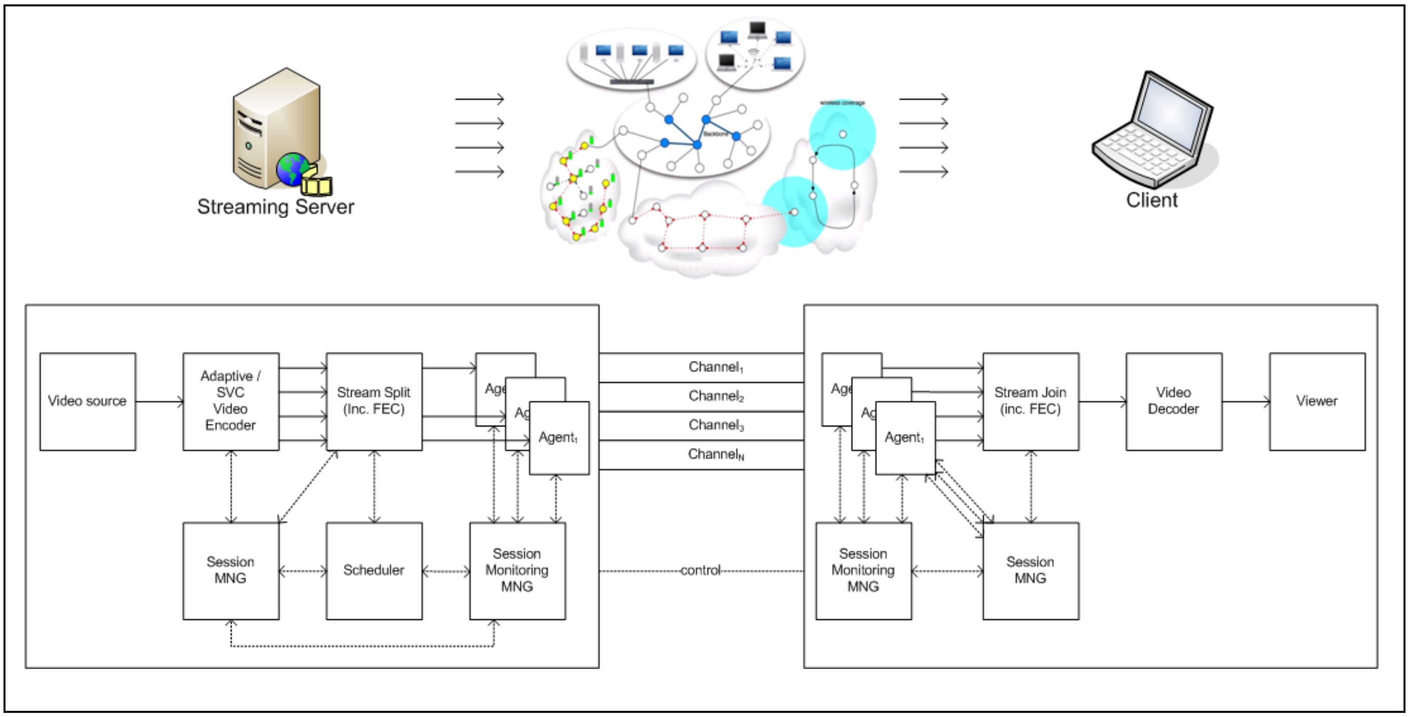


Figure 1. Architecture of the multi-channel video streaming system [3]

II. Multi-Channel Video System

In this section, we provide an overview of the multi-channel video transmission system under consideration [3]. As shown in Figure 1, the system consists of three parts: the video server, the multiple channels and the video client. The server and clients can communicate over up to M multiple channels (paths, networks). The video server consists of a video source, a video encoder, a module for stream splitting and channel protection, a session monitoring module, channel scheduler and a session manager module. The video client consists of a module for joining and decoding the channel protection, a session monitoring manager module, a session manager module, a video decoder and a viewer. These components are described briefly in the following paragraphs.

We assume that a space-time discrete video signal is used as input to the layered video encoder, which is characterized by its operational distortion-rate function. After source coding, the compressed layered video stream is prepared for transmission by the channel codec. This involves packetization and Forward Error Correction (FEC) combined with interleaving to reduce the effect of burst errors. After channel encoding, the video layers are scheduled to active channels and then the video packets are transmitted over the channels according to their layer to channel mapping.

In a general multi-channel network, different channels may have different parameter values. Furthermore, channel parameters may change due to the activation of other channels that share some resources, such as bottlenecked links [22]. We used the enhancement of the Gilbert-Elliott model [1][2] into a packet erasure multi-channel model [22] to characterize the multi-channel behavior in terms of video rate and error rate. According to this model, the multi-channel video rate for homogeneous channels is generated using the following formula:

One channel: $R_1 = R \cdot (1-\alpha_1)$

Two channels $R_2 = 2R \cdot (1-\alpha_2)$

$$\cdots$$

M channels $R_M = MR \cdot (1-\alpha_M)$

Where $\alpha_1 < \alpha_2 < \cdots < \alpha_M$. That is, the error rate increases with the number of active channels [22].

III. Algorithms for Session Management of Multi-Channel Variable Bit Rate Video Streaming Session

The session manager's tasks are:

1) *Calculate: target video rate and additional parameters.*
2) *Decide: the number of active channels A* $(1 \le A \le M)$

In this paper we focus on the question of deciding the number of active channels.

In this section, we propose two session management algorithms in a multi-channel video system for transmitting video with variable video rate. The first algorithm is a simple algorithm and the second one is a feedback-based algorithm. These algorithms are compared with each other and with static sessions (unmanaged), in which the number of active channels is constant.

The simple session management procedure is as follows.

```
Void SimpleSessionMNG_Procedure()
Begin
1:  In the first time interval Do:
1.1:   Initiate session;
1.2:   Activate all M channels;
2:  In the second time interval Do:
2.1:   Get report from session monitor;
3:  For each time interval i>1 Do:
3.1:   Calculate target video rate $R_V$;
3.2:   Activate A+1 channels such that $R_A \geq R_V$;
3.3:   Update the other modules;
End
```

The simple-managed module initiates the session. In the first time interval, the algorithm activates all *M* channels. The algorithm gets a measurement report only once (after the first time interval), and learns the effective bandwidth of each channel from this report. In the following time intervals, the algorithm decides the number of active channels according to the target video rate of each time interval and the effective bandwidth of the channels that was reported at the beginning of the session. The channels are chosen sequentially. Either by index number (arbitrarily defined), or by descending order of bandwidth, that was estimated based on the report of the first interval.

The feedback-based session management procedure is as follows.

```
Void FeedbackSessionMNG_Procedure()
Begin
1:  In the first time interval Do:
1.1:   Initiate session;
1.2:   Activate all M channels
2:  For each time interval i>1 Do:
2.1:   Get report from session monitor;
2.2:   Calculate target video rate $R_V$;
2.3:   Decide the number of active channels A;
2.4:   Update the other modules;
End
```

The feedback-managed module also initiates the session and activates all *M* channels in the first time interval. Afterwards, in each time interval, the algorithm calculates the target video rate, decides the number of active channels and chooses the particular set of channels. Finally, it updates the other modules.

The next procedure describes the algorithm for deciding the required number of active channels according to session history and the target video rate in the next time interval (step 2.3 of the feedback-managed procedure).

```
Int FeedbackManagedActiveChannels ($R_V$, M,
                     timeInterval, channelSize[])
Begin
1:  Calculate last interval arrival percent
       based on monitor report;
2:  For each channel i in M channels Do:
2.1:  set availableChannelSize[i] =
         channelSize[i]*lastIntervalPercent[i];
3:  If $\sum^{M} availableChannelSize < R_V$
3.1:  Return M;
4:  Find min A such that
       $\sum^{A} availableChannelSize \geq R_V$
5:  Return A+1;
End
```

In each time interval, the algorithm gets a measurement report of the active channels from the session monitoring manager and calculates the percent of the data that arrived on time out of the sent data. Then the current available bandwidth for real-time transmission of each channel is determined. Finally, the algorithm finds the minimum number of channels whose available bandwidth sum is sufficient for the target video rate in the next time interval.

The task of trying to find the minimum number of channels that satisfies the video rate (step 4 of the FeedbackManagedActiveChannels procedure) can be performed in several ways that differ by complexity and accuracy. Different methods can provide different results. We suggest two possible methods: a simple sequential method and a more complex optimal method. The first method is described in the next procedure.

```
Int FindSequentA_Procedure($R_V$)
Begin
1:  For k=1 to M
1.1:  If $\sum_{i=1}^{k} availableChannelSize \geq R_V$
1.1.1:     return k;
End
```

The FindSequentA procedure chooses the channels sequentially. In each iteration, the procedure tests the channels indexed 1 to k. If the sum of their available size is enough, the first k channels are chosen. The sequential method is very simple to compute, but does not always provide the best minimal set of channels. The second suggested method is described below:

```
Int FindOptimalA_Procedure()
Begin
1: For n=1 to M
1.1:   Find subset N of n channels such that:
         $\sum^{N} availableChannelSize$ Is maximal;
1.2:   If $\sum^{N} availableChannelSize \geq R_V$
1.2.1:         return n;
End
```

The FindOptimalA procedure uses exhaustive search to choose the set of *A* channels as finding such subset is NP-complete [29]. In each iteration, the procedure tests all $\binom{M}{n}$ subsets of *n* out of the *M* possible channels, 1≤n≤M, and finds the set with maximal bandwidth. If that set (with maximum available channels size) has enough bandwidth, that set is chosen. The method gives optimal results by examining sets of channels of varying sizes. It starts with sets of one channel, then tests sets of two channels until the set of *M* channels. When the procedure finds a sufficient set – that set is chosen. This order of testing the sets ensures the optimal choice with minimum number of channels and maximum bandwidth.

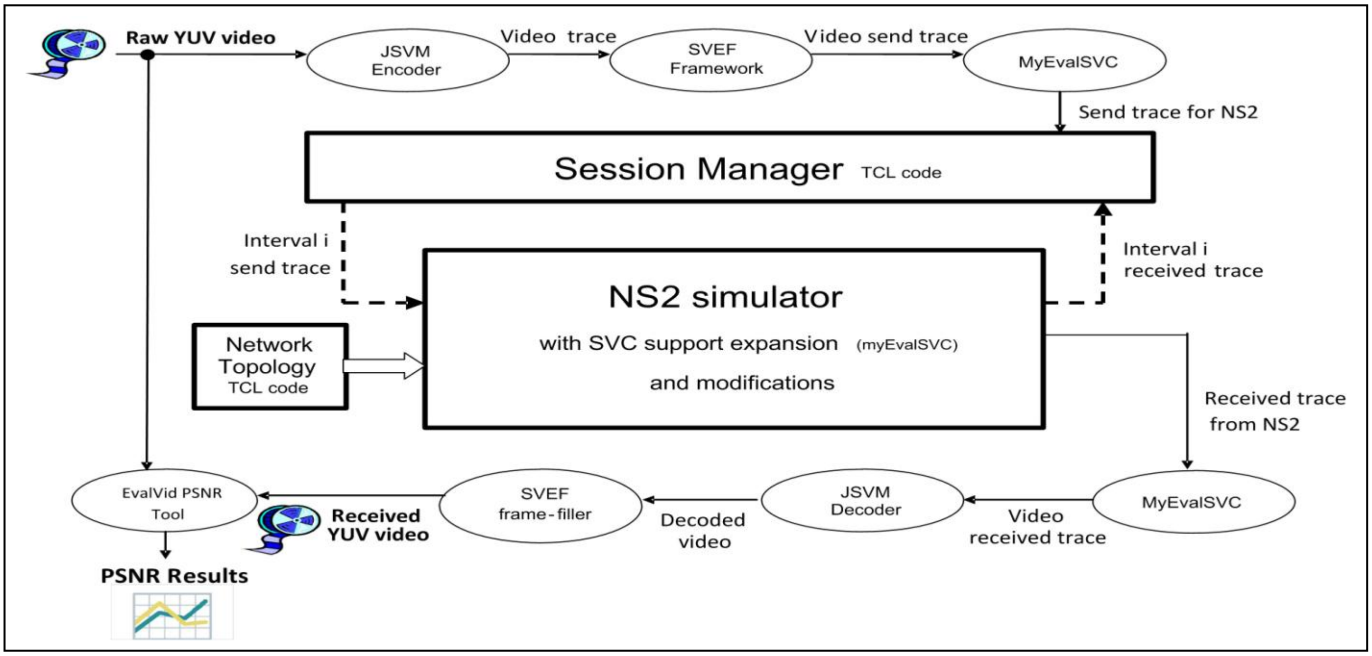


Figure 2. The Simulation Environment

IV. Implementation With Network Simulator

In this section, we describe the simulation environment used for evaluating the suggested session management algorithms performance. The environment is Network Simulation – 2 (NS-2) based. NS-2 is an open-source network simulator widely used in academic research [23]. The simulator is fed with scripts and traces to be sent over the simulated network. It simulates the network behavior and generates traces for the receiving end. We implemented the session-manager algorithm in TCL and ran it on NS-2. In addition, we used the following open-source tools:

- The JSVM Reference Software [24] - the reference software for the Scalable Video Coding (SVC) project of the Joint Video Team (JVT).
- The SVEF Framework [25] – a Scalable Video coding streaming Evaluation Framework, devised to evaluate the performance of H.264 SVC video streaming.
- The MyEvalSVC [26] - an integrated simulation framework for evaluation of SVC transmission based on the SVEF and extended to connect to the NS2 simulator.
- The EvalVid [27] - a tool-set developed for evaluation of the quality of a video transmitted over a real or simulated network.

In Figure 2, the simulation environment is presented. As can be seen from this figure, a raw YUV video is encoded using the JSVM-Encoder and a video trace file is created by JSVM's Bit Stream Extractor. The trace file is processed in the SVEF's f-nstamp tool and a send-trace file is generated. The send-trace is converted to the format of a NS2-send-trace file using MyEvalSVC tools. The full NS2-send-trace is delivered to the session manager module. In each time interval (one second of video), the manager splits the interval's part of the NS2-send-trace into *A* different NS2-send-traces, according to the algorithm decision of *A* active channels. Then, NS-2 simulates sending the video interval over *A* channels out of the *M* channels we defined in the network topology. When the interval is completely received on the receiver end, the session manager module gets the received-trace, and based on the performance that is derived from the send-trace and received-trace, it decides the number of active channels for the next interval. The manager repeats this routine until the entire video is transmitted.

When the simulation is completed, the full NS2-received-trace of the video is converted into a video trace using MyEvalSVC tools. The JSVM-Decoder decodes the received video trace back to a raw YUV video file. Since some frames might be lost during the transmission, the video is reconstructed using the SVEF frame-filler tool, which fills missing frames by duplicating other frames, so that the received video has the same length as the sent video. Finally, the quality of the received video is evaluated according to its frames PSNR value, which is calculated by comparing it to the original video with EvalVid's PSNR tool.

V. Simulation Results

In this section, we describe the benchmark, the simulated network topology, and the performance of the session management algorithms.

To validate the session manager’s performance, we decided to use the PSNR matric that measures video quality. The PSNR value is calculated by comparing two raw YUV formatted video sequences. Therefore, we searched for long original YUV sequences. Most of the video traces include short video only, and were not suitable to study the management algorithms benefit and limitations. The most suitable sequence we found was the open source animation video Big Buck Bunny [28], a long high-quality YUV sequence. We simulated transmitting 5 minutes of the video. The frame rate of the video was 24 fps and we defined time interval length to be one second of video. The total number of transmitted intervals was 300 and the number of frames was 7200.

The transmission of the video was tested over two different topologies. The first topology was very simple. The source and the destination were connected with five independent channels. In this case, there was no significant difference between the different management algorithms. In reality, in a multi-channel network, independent channels are very uncommon. Hence, we tested the second network topology as presented in Figure 3 that better reflects realistic multi-channel conditions. In this network topology, the source is connected to the destination with five channels as well. But, the first two channels are occasionally disturbed by other entities that use their resources (source 2 transmits to destination 2 and source 3 transmits to destination 3) and channels three and four correspond to two edge dependent paths. The fifth channel is independent. The propagation delay of each link is 1ms, and the bandwidth is 4Mbps.

We simulated both simple and feedback managed algorithms, each one with the two methods of channels selection we described. The results of both versions of the simple manager were very similar, hence we will present only the results of one method for the simple manager. We assume that sorting the channels (the second method) does not improve the results compared to sequentially choosing the channels (the first method), because the simple manager gets report only after the first interval, and that report does not sufficiently reflect the channels' capacity that changes over time.

Figure 4, 5, and 6 present the simulation results of the feedback-managed and simple-managed algorithms. Throughout this section, we refer to *information loss* in case of either actual information loss or in case of late arrival (according to the video play-time). Figure 4 presents the results of the byte-loss percent (a), and the video frame-loss percent (b) for each time interval for the optimal and sequential feedback-managed algorithms (green and red lines) and simple-managed algorithm (blue line). We can see in both graphs that the optimal feedback-managed algorithm lost the least amount of data, the sequential feedback-managed algorithm lost more data than the optimal and the simple-managed algorithm lost the most amount of data.

We refer to the loss of data in two aspects: byte-loss and video frame-loss. The difference between the two aspects is due to the simple open source decoder that decodes the received video. The decoder is not sophisticated enough to handle decoding of partial frames, therefore even if only a few bytes are missing, a whole frame is deleted and possibly other frames depending on this frame. For example, in time interval number 229 (marked with a dashed line number 1 in Figure 4), the difference is noticeable. The red graph that represents the loss of the sequential feedback-managed algorithm in very low (0.9%) in graph (a) (byte-loss), but is significantly higher (9 out of 24 video frames) in graph (b) (frame-loss). An additional example that emphasizes the difference is provided in time interval 35 (marked with dashed line number 2). In graph (b), the red and blue graphs (sequential feedback-managed and simple-managed respectively) are very close, meaning they lost almost the same number of video frames (11 lost video frames using the feedback-managed algorithm vs. 12 lost video frames using the simple-managed algorithm). However in graph (a) at the same point, the difference between the two graphs is distinguishable (17% byte loss vs. 33% byte loss).

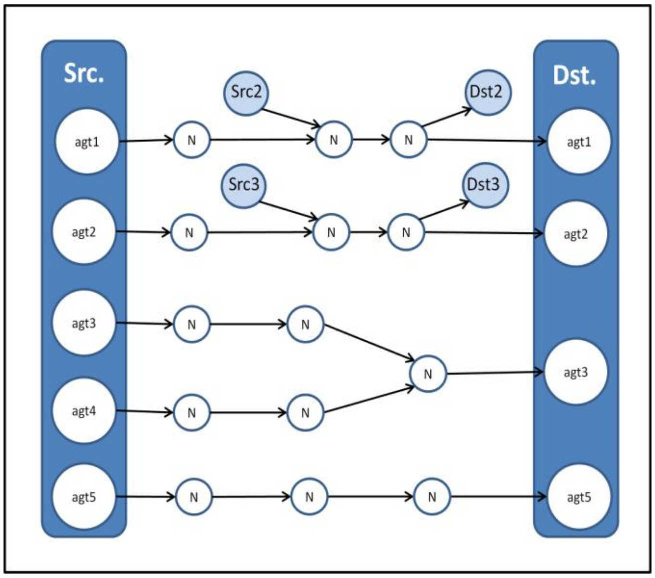
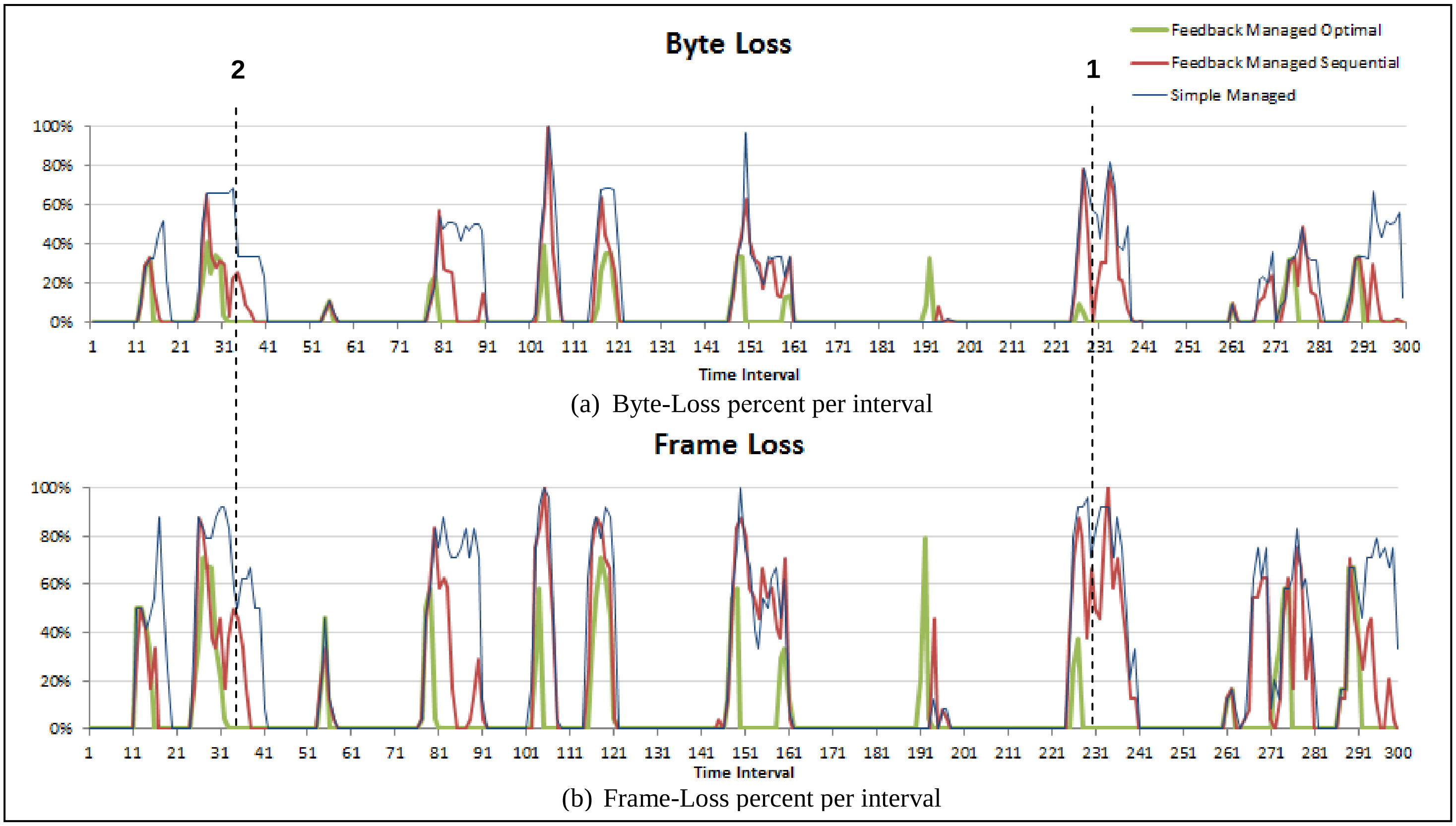


Figure 3. Simulated network topology

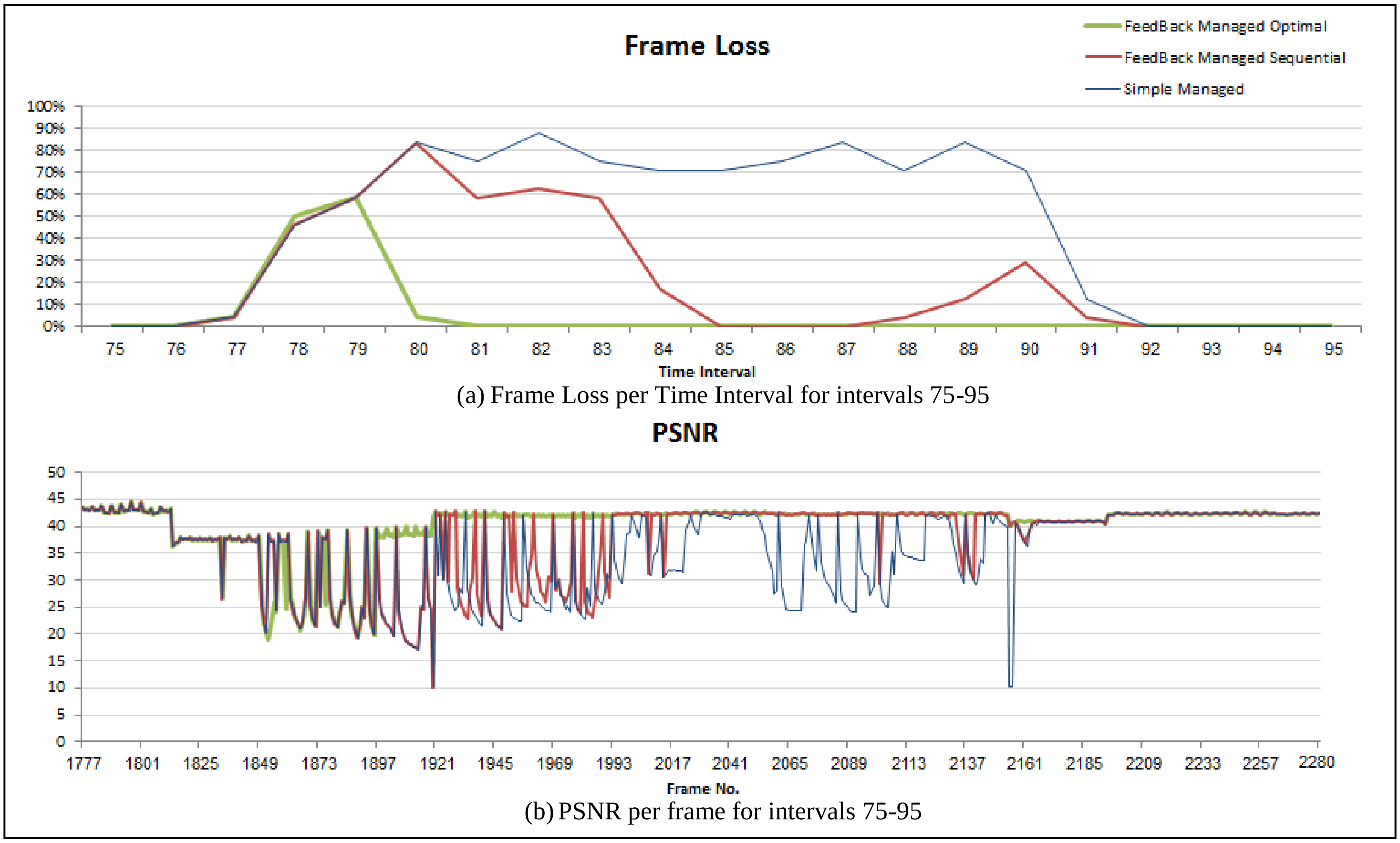
Figure 5 provides an example of the performance results with respect to the PSNR. The results of frame-loss (a) and PSNR value (b) for the two feedback-managed and simple-managed algorithms in time intervals 75-95 are plotted in this figure. The frame loss was calculated per interval, while the PSNR was calculated per frame, so that each point in graph (a) that represents one interval is equivalent to a sequence of 24 frames in graph (b). Clearly, high video frame loss will cause low PSNR. The PSNR value is calculated by comparing the original YUV sequence to the YUV sequence that was decoded from the received SVC video. The PSNR value is affected by two factors. First, the encoding and compressing of the original video by the video encoder before it was sent over the simulated network (from YUV to SVC). The second factor is the data loss caused by the video transmission over unreliable channels. All algorithms were affected identically by the first factor. The difference in the PSNR between them was caused only by the second factor, hence, the difference between the graphs points out the advantage of both of the feedback-managed algorithms over the simple-managed algorithm, and the advantage of the optimal method over the sequential method. In time intervals 75-76, 92-95, there was no loss of data in all algorithms, as can be seen in graph (a), and the PSNR value (shown in graph (b)) is high because it was affected only by the encoding process. In time intervals 77-80, the same number of frames were lost in the sequential feedback and the simple algorithms and the PSNR value is also identical, but the optimal feedback managed lost a lot less frames in intervals 79-80 and its PSNR value is much higher. In the rest



(a) Byte-Loss percent per interval

(b) Frame-Loss percent per interval

Figure 4. (a) Byte-Loss; (b) Video Frame-Loss for simple-managed (blue line), optimal (green line) and sequential (red line) feedback-managed algorithms



(a) Frame Loss per Time Interval for intervals 75-95

(b) PSNR per frame for intervals 75-95

Figure 5. (a) Video Frame Loss; (b) PSNR for intervals 75-95 for simple-managed (blue), optimal (green) and sequential (red) feedback-managed algorithms

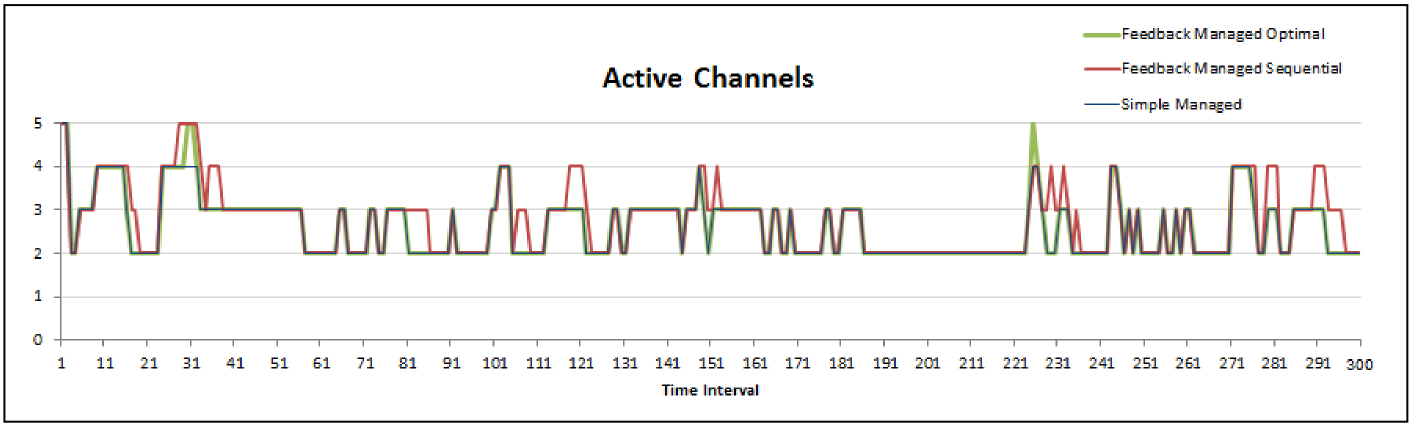


Figure 6. Number of active channels per time interval for simple-managed (blue), optimal (green) and sequential (red) feedback-managed algorithms

of the time intervals the optimal feedback-managed (green line) did not lose any frames and its PSNR value is the highest. In time intervals 81-84, graph (a) shows that the sequential feedback manager (red line) lost less video frames than the simple-manager (blue line), and graph (b) also shows that the PSNR of the sequential feedback managed is slightly higher than the simple-managed. In time intervals 85-91, the recovery of the sequential feedback-managed is clearly seen, in graph (a) its frame loss is very low and in (b), its PSNR is high, in contrast to the simple-managed algorithm whose frame loss remains high, and PSNR remains low.

Figure 6 plots the number of active channels during the simulation period. Due to the changes in the channels conditions that are reported only to the feedback-managed algorithms, there are intervals where the feedback-managed algorithms activates more channels than the simple-managed algorithm, as seen in the graph. The graph also shows that frequently the optimal feedback manager (green line) activates less channels than the sequential feedback manager (red line), because it chooses the channels with the biggest capacity, and therefore is satisfied with fewer channels.

TABLE I. PERFORMANCE EVALUATION

Method	Byte Loss	Active Channels Avg.	PSNR		PSNR violation
			Mean	Std.	
Feedback-Managed-OptimalA	5.13%	2.6	40.01	4.77	12.15%
Feedback-Managed-Sequential	10.72%	2.74	38.51	6.74	20.86%
Simple-Managed-Bandwidth-order	18.09%	2.59	37.35	7.82	27.34%
Simple-Managed-Sequential	17.68%	2.59	37.32	7.83	27.52%
Static 2-channels	46.72%	2	31.84	11.82	48.5%
Static 3-channels	15.44%	3	38.14	7.06	22.95%
Static 4-channels	11.89%	4	38.9	6.36	18.37%
Static 5-channels	4.92%	5	40.04	4.74	11.22%

Table I summarizes the performance evaluation of the two suggested session management algorithms and static sessions with 2,3,4,5 channels. A Static i-channels method (lines 3-6 in Table I) is a simple un-managed method in which the first i'th channels are selected to transmit the video regardless of the target video rate and the channels temporal conditions. For each method, we present in the table the results of: (i) Byte loss percent – the percent of the data that was lost in the transmission or arrived too late (considering a 1000ms play-out buffer). (ii) Active channels average – the average number of channels that were activated per time interval. (iii) PSNR – the average value of PSNR per frame and the standard deviation of PSNR per frame. The PSNR average alone does not fully reflect the quality of the received video, because if the average is high but the standard deviation is also high, the viewing experience is impaired due to the significant changes in the PSNR value between frames. (iv) PSNR violation – the percent of frames whose PSNR value is lower than 37, which represents a minimum threshold for high video quality according to MOS mapping.

As expected, the static 5-channels method outperforms the other method with high average PSNR and low PSNR std. However, since it activates five channels, it has the highest cost (assuming that every channel has its own operation costs). The static 2-channels method has the lowest performance and the lowest cost. Almost half of the information did not reach the destination on time, and it violates the top MOS PSNR level on 48.5% percent of the video frames. Between these performance edges, we have the static 3-channels, 4-channels and the two managed methods. It can be seen that the feedback-manage method successfully balances between cost, on-time information delivery and PSNR results. Note that although the static 4-channels method outperforms the feedback-managed method with slightly higher PSNR and slightly lower PSNR violation, the feedback-managed, in fact, delivered more information to the destination on-time and the gap is a result of the simple decoder in use in the simulation only.

VI. CONCLUSION

In this study we suggested two simple algorithms for session management in multi-channel variable bit rate video transmission. The algorithms have the following properties: they are very simple to compute, they require low computation afford, low storage requirements, and small feedback messages that provide the statistics for each active channel in the previous time interval (used only by the second algorithm).

The evaluation of the management algorithms is based on simulation environment. The simulation results show that it is possible to control the session costs in terms of the number of active channels while keeping the quality of the received video stream in the top MOS level. For example, compared to the static selection of three channels, the simple management algorithm achieved a 13.67% percent cost reduction with 2.59 active channels on average, and average PSNR of 37.32 compared to the average PSNR of 38.14 (reduction of only 2.15%). The feedback-based management algorithm achieved an 8.67% percent cost reduction with 2.74 active channels on average. Furthermore, the feedback-managed algorithm delivered 89.28% bytes on-time compared to only 84.66% bytes that were delivered on-time using a static selection of three channels. This leads to an average PSNR of 38.51, that is, a 0.97% percent improvement in the average PSNR compared with a static selection of three channels. Using a more sophisticated decoder could improve the PSNR even more.

ACKNOWLEDGMENT

We thank Net-HD video team and in particular Noam Amram for useful discussions. We thank Mordechai Behar for technically supporting this research.

REFERENCES

[1] E. O. Elliott, “Estimates of Error Rates for Codes on Burst-Noise Channels”, Bell System Technical Journal 42, 1963, pp. 1977–1997.

[2] E. N. Gilbert, “Capacity of a Burst-Noise Channel”, Bell System Technical Journal 39, 1960, pp. 1253–1265.

[3] R.Nossenson and N. Amram, "Session Management in a Multi-Channel Video Streaming System for Wireless Networks", The 18th IEEE International Conference on Networks (ICON 2012), Singapore, Dec. 2012, pp. 333-338.

[4] R.Nossenson and N. Amram, "Packet Scheduling in Multi-Channel Layered-Video Streaming over Wireless Sensor Networks", To appear, The 1st International Workshop on Wireless Multimedia Sensor Networks (WMSN'12) (part of WiMob 2012), Barcelona, Spain, Oct. 2012, pp. 683-688.

[5] R. Nossenson and O. Markowitz, “Using Coordinated Agents to Improve Live Media Content Transmission”, In proceedings of the Sixth International Conference on Systems and Networks Communications (ICSNC 2011) , Barcelona, Spain Oct. 2011, pp. 167-170.

[6] J. Apostolopoulos, T. Wong, W. Tan, and S. Wee. "On multiple description streaming with content delivery networks", In IEEE INFOCOM, 2002, pp. 1736-1745.

[7] E. Setton, X. Zhu and B. Girod, “Minimizing distortion for multipath video streaming over ad hoc networks”, IEEE Int. Conf. Image Processing (ICIP-04), Singapore, vol 3, Oct. 2004, pp.1751–1754.

[8] S. Mao, S. Lin, S. S. Panwar, Y. Wang, and E. Celebi, "Video Transport Over Ad Hoc Networks: Multistream Coding With Multipath Transport", IEEE journal on selected areas in communications, Vol. 21, No. 10, Dec. 2003, pp. 1721-1737

[9] B. Wang, W. Wei, and D. Towsley, "Multipath Live Streaming via TCP: Scheme, Performance and Benefits", ACM Transactions on Multimedia, 2009, pp. 11.1-11.12.

[10] D. Jurca and P. Frossard, “Distributed media rate allocation in multipath networks”, Signal Process.: Image Commun., Vol.23, 2008, pp.754–768.

[11] J. Apostolopoulos, T. Wong, W. Tan, and S. Wee. "On multiple description streaming with content delivery networks", In IEEE INFOCOM, 2002, pp. 1736-1745.

[12] L. Golubchik et al, “Multi-path continuous media streaming: What are the benefits?”, Performance Evaluation, 2002, pp. 429-449.

[13] Y. J. Liang, E. G. Steinbach, and B. Girod, “Real-time voice communication over the Internet using packet path diversity”, In ACM Multimedia, Ottawa, Canada, Sep. 2001, pp. 431-440.

[14] T. P. Nguyen, and A. Avideh, “Mutiple sender distributed video streaming”, IEEE Transaction on Multimedia, 6(2), April 2004, pp. 315-326.

[15] M. Wang, L. Xu, and B. Ramamurthy, "Linear Programming Models For Multi-Channel P2P Streaming Systems", Mini-Conference at IEEE INFOCOM 2010, pp. 1-5.

[16] E. S. Ryu, H. Kim, S. Park, and C. Yoo, "Priority-Based Selective H.264 SVC Video Streaming Over Erroneous Multiple Networks", 2011 IEEE International Conference on Consumer Electronics (ICCE), 2011, pp. 337-338.

[17] L. Zhou, X. Wang, W. Tu, G. M. Muntean, and B. Geller, "Distributed Scheduling Scheme for Video Streaming over Multi-Channel Multi-Radio Multi-Hop Wireless Networks", IEEE journal on selected areas in communications, Vol. 28, No. 3, April 2010, pp. 409-419.

[18] L. Zhou, B. Geller, B. Zheng, S. Tang, J. Cui, and D. Zhang, "Distributed Scheduling for Video Streaming over Multi-Channel Multi-Radio Multi-Hop Wireless Networks", in proceedings of IEEE GLOBECOM 2009, pp. 409-419.

[19] V. Agarwal, and R. Rejaie, “Adaptive Multi-Source Streaming in Heterogeneous Peer-to-Peer Networks,” Proc. Multimedia Computing and Networking (MMCN ’05), Jan. 2005, pp. 13-25.

[20] Y. Guo, C. Liang, and Y. Liu, “AQCS: Adaptive Queue-based Chunk Scheduling for P2P Live Streaming,” in Proceedings of IFIP Networking, 2008, pp. 433-444.

[21] M. Zhang, Y. Xiong, Q. Zhang, and S. Yang, “Optimizing the throughput of data-driven peer-to-peer streaming”, IEEE Transactions on Parallel and Distributed Systems, vol.20, no.1, 2009, pp. 97-110.

[22] R. Nossenson, and N. Nossenson, “Packet Erasure Model for Multi-Channel Video Streaming”, to apear, IEEE International Symposium on Network Computing and Applications (NCA), 2015.

[23] NS2 Simulator http://www.isi.edu/nsnam/ns/ , retrieved Sep. 2015.

[24] JSVM Software, http://evalsvc.googlecode.com/files/SoftwareManual.doc, retrieved Sep. 2015.

[25] SVEF Fraemwork, http://svef.netgroup.uniroma2.it/ retrieved Sep. 2015.

[26] MyEvalSVC Toolset, https://code.google.com/p/evalsvc/ retrieved Sep. 2015.

[27] EvalVid Toolset, http://www.tkn.tu-berlin.de/menue/research/evalvid/ retrieved Sep. 2015.

[28] BigBuckBunny YUV Sequence, http://www.bigbuckbunny.org/ retrieved Sep. 2015.

[29] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, “Introduction to Algorithms” (3rd ed.). McGraw-Hill Higher Education, 2009.

Optical Last Miles for Research and Education in the Czech Republic

Jan Radil, Lada Altmannová, Ondřej Havliš, Miloslav Hůla, Stanislav Šíma, Josef Vojtěch
Optical Networks Department
CESNET, Association of Legal Entities
Praha, Czech Republic
jan.radil@cesnet.cz, lada.altmanova@cesnet.cz, ondrej.havlis@cesnet.cz, miloslav.hula@cesnet.cz,
stanislav.sima@cesnet.cz, josef.vojtech@cesnet.cz

***Abstract* — There are compelling reasons for building networks which are future-proof, scalable, and which will be able to accommodate power users with special needs in the future. However, while the backbone networks are typically ready for new trends such as coherent systems, the situation with access networks is different, often for economic reasons. In this article, we present a cost-effective solution based on open equipment which can be advantageous even outside the academia, research and education ecosystems. As an example, we describe our use of the Czech Light® family of devices within the central region of the CESNET's production network.**

Keywords – optical fiber network; metropolitan area network.

I. Introduction

National research and education networks (NREN) provide connectivity for universities, research centres and other advanced users. Their backbone networks use dense wavelength division multiplexing (DWDM) coherent transmission systems rather routinely, and the data rates of 100 Gb/s are quite common. Successful 1 Tb/s trials have been performed in networks where dark fibers are abundant and available for experiments [1] [2].

This transmission capacity situation is rather different in access parts of the networks. The DWDM systems are not deployed commonly and transmission speeds are usually limited to 10 Gb/s. Sometimes the legacy time division multiplexing (TDM) technology is still used. This stark contrast with the backbone networks is often caused by economic reasons as upgrades are not conducted that often.

Moreover, there are certain new scientific applications with rather special requirements. Examples of these are an accurate time transfer, or a very stable frequency transfer. For these applications, increasing the transmission speeds to 100 Gb/s or even 1 Tb/s are not important and will not help when such applications are deployed [3]. The reason for this constraint is that the time and/or frequency transfer is not about ‘big data’ transfers, but rather about stable and very low jitter. An accurate time transfer uses speeds well below 1 Gb/s. A transmission of stable frequency consists of a so called continuous wave (CW) signal, i.e., a signal without any modulation because the frequency of photons is the useful property of the transmitted information.

Very accurate time and ultra-stable frequency are crucial for many fields, for example sensing, metrology, navigation, geodesy, radio-astronomy, Earth surveying, seismology, fundamental physics, etc. The increased interest in the all-optical time and frequency transfers are manifested by the EU joint research project NEAT-FT [4].

Unfortunately, technical issues may arise when high speed coherent systems and time/frequency applications are operated together over a shared fiber infrastructure with regular data over DWDM [5].

In this contribution, we will describe some practical examples that show how CESNET has been able to overcome these economical and technical issues.

II. Optical Last Miles Issues

The issues related to last miles are well-known and all operators have learned to deal with them. Sometimes Last Miles have been dubbed as First Miles to emphasize their importance for high speed optical networking. In an NREN ecosystem, the last miles' problems cannot be solved by means of wireless networking because capacity (or bit rate) is not large enough for big demanding applications. With the higher bit rates, one has to utilize higher carrier frequencies, but their reach decreases significantly.

One example of such demanding application can be the ultra-high definition video transmission required for medical applications [6]. Moreover, new applications such as hard real-time controls require very low and constant jitter which can be satisfied successfully with an optical fiber [7]. Real-time network services are needed for an interaction with external processes, in other words for any processes running outside the network. Examples of these use cases include collecting data from remote sensors or telescopes, or remote machine control. The importance of these topics can be found, for example, in the Strategy document for the pan-European network GÉANT for the 2020 time frame [8].

To provide new opportunities for the research, education and scientific community, CESNET has developed new equipment – the Czech Light® family of advanced photonic devices. All of the Czech Light® devices are open. The word ‘open’ means that third parties are allowed to modify the Czech Light® devices, so it is easy to deploy them in new networking scenarios. The Czech Light® devices can be also customized by power end users, e.g., by augmenting them with a custom, specific control software. This is usually not possible with equipment from traditional big vendors.

III. CESNET Solutions for the Last Miles

Dark fibers have been used in the CESNET network for many years. The first dark fiber was lit back in 1999, with Packet over SONET (PoS) technology with 2.5 Gb/s speed. At that time electro-optical regenerators for SONET/SDH were the primary option for extending the reach. Later on, optical amplifiers started to emerge, especially when optical gigabit Ethernet was deployed in metropolitan (MAN) and even wide area networks (WAN).

The Czech Light® optical amplifiers (CLA) have been developed by CESNET to overcome limitations of the then-available optical equipment. The most significant drawback of contemporary commercial offerings was the lack of standardized monitoring capabilities. Support for the de-facto standard Simple Network Monitoring Protocol (SNMP) was one of the key requirements for practical deployment for any NREN or Internet Service Provider (ISP).

The Czech Light® amplifiers are based on commercially available modules of Erbium doped fiber amplifier (EDFA). The Czech Light® family of devices also include reconfigurable add-drop multiplexers (ROADM), wavelength selective switches (WSS) or tuneable dispersion compensators (TDC). All of these devices consist of the optical module, an embedded Linux system, and essential control electronics. The Czech Light® devices are housed in a standard rack chassis of size 1U or 2U, and can be customized on demand.

Various Czech Light® products are protected by several patents in the EU [9] and within the US [10] [11] [12].As of 2015, the Czech Light® equipment is used on 4890 km of the CESNET networks, including 2012 km of bidirectional single-fiber transmission.

Figures 1 to 4 show the up-to-date situation with Czech Light® equipment deployed in some of the important optical last miles in the central area around Prague (Praha).

Fig. 1 shows a bidirectional single fiber line between Praha and Dolní Břežany, where the Extreme Light Infrastructure (ELI) is located. This design features so called one-side amplification, where an active device is located at one end of a fiber line only. In this case it is one optical amplifier Czech Light® with two independent modules, one serving as a power amplifier (booster) and another one serving as a preamplifier.

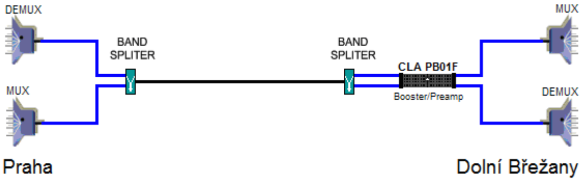


Figure 1. Bidirectional single fiber line Praha-Dolní Břežany.

Fig. 2 shows a standard fiber line between Praha and Řež, a site of numerous research centres and institutions.

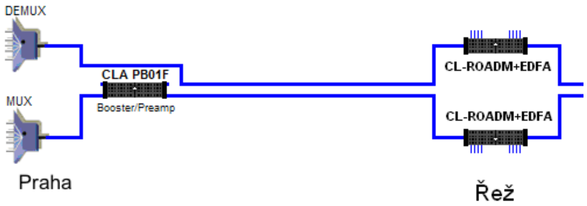


Figure 2. Standard fiber line Praha-Řež.

Fig. 3 shows the ring Praha-Řež-Jenštejn-Praha which is used to increase reliability of a critical part of the network. Both remote locations are hereby reachable from two geographically different directions. Within this path, only the Řež-Jenštejn segment is built on a bidirectional single-fiber line. Both end of the fiber in Praha terminate at the same physical location.

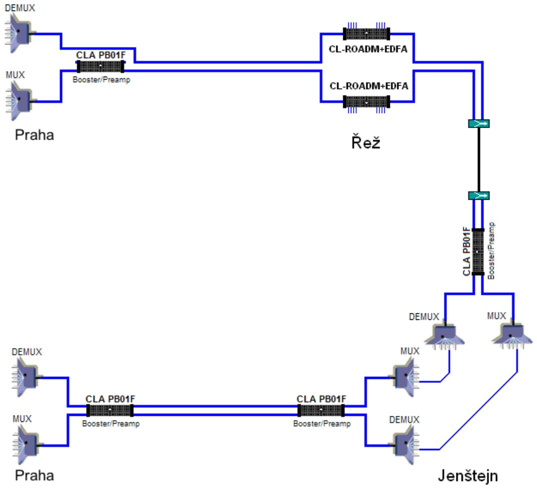


Figure 3. The ring Praha-Řež-Jenštejn-Praha.

Fig. 4 shows the ring Praha-Krč-Vestec-Dolní Břežany-Praha for ELI and BIOCEV. Vestec hosts the Biotechnology and Biomedicine Centre (BIOCEV) of the Academy of Sciences and Charles University. The entire ring is built upon bidirectional single-fiber segments. Both ends of the fiber ring in Praha terminate at the same physical location.

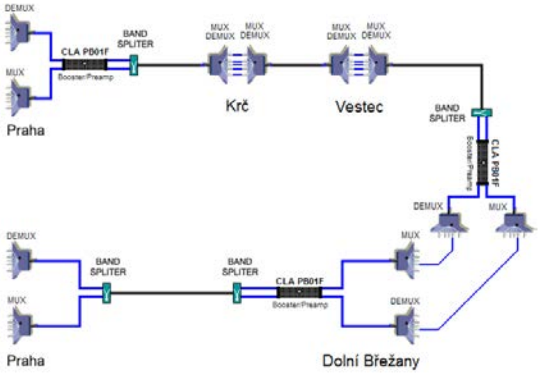


Figure 4. The ring Praha-Krč-Vestec-Dolní Břežany-Praha.

Figures 3 and 4 also demonstrate the deployment of multiple Czech Light® ROADMs. These optical last miles based on the Czech Light® equipment therefore support dynamic, on-demand reconfiguration in response to a remote command. This feature allows CESNET to better optimize usage of its fiber pool as future requests for circuits arrive. As an additional reference, Fig. 5 shows the schematic topological runs of the two described dark fiber rings, with a major part of their length being deployed over bidirectional single fiber lines.

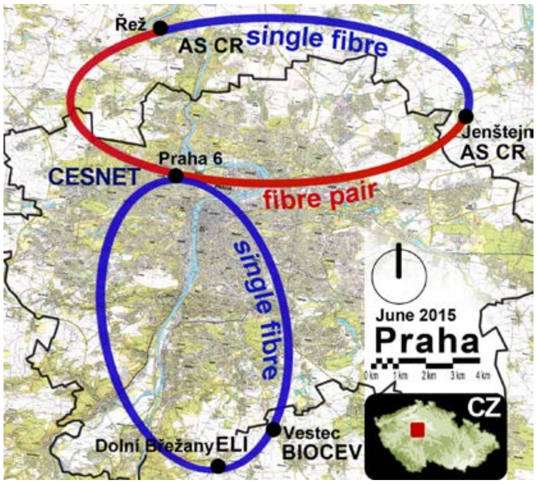


Figure 5. Schematic topology of dark fiber rings.

IV. CONCLUSION

In this paper, we presented the feasible technical solutions of the optical last miles used in the Czech NREN CESNET. The added value of the CESNET solutions is twofold: economic feasibility allowing for higher transmission capacity and openness allowing for deployment of new applications without any restrictions.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education, Youth and Sport of the Czech Republic as part of the CESNET Large Infrastructure project LM2010005.

REFERENCES

[1] Infinera and DANTE Demo Single-Card Terabit Super-Channel. [Online]. Available from: http://www.convergedigest.com/2014/09/infinera-and-dante-demo-single-card.html 2015.08.26

[2] Tera Santa Consortium Demos 1 Tbps with ECI, Finisar, Technion. [Online]. Available from: http://www.convergedigest.com/2014/04/tera-santa-consortium-demos-1-tbps-with.html 2015.08.26

[3] J. Vojtěch, V. Smotlacha, P. Škoda, "Optical infrastructure for precise time and stable frequency transfer," Proc. SPIE 8866, Earth Observing Systems XVIII, 09/2013, pp. 1-5, doi:10.1117/12.2024405.

[4] Accurate Time/Frequency Comparison and Dissemination through Optical Telecommunication Networks. [Online]. Available from: https://www.ptb.de/emrp/neatft_home.html 2015.08.26

[5] L. Altmannová et al., "Photonic Services: Challenge for Users and for Networkers", GN3Report, 2013, pp. 1-13. [Online]. Available from: http://geant3.archive.geant.net/Media_Centre/Media_Library/Media%20Library/GN3-13-110_JRA1-T2_Photonic-Services.pdf 2015.08.26

[6] CESNET got attention with medical 4K streaming in Denver. [Online]. Available from: http://www.cesnet.cz/cesnet/reports/press-releases/cesnet-got-attention-with-medical-4k-streaming-in-denver/?lang=en 2015.08.26

[7] Šíma, S., "New requirements for R&E networks",7th Customer Empowered Networks workshop, 2012. [Online]. Available from: archiv.ces.net/events/2012/cef/p/New%20requirements%20for%20R&E%20networks.pdf 2015.08.26

[8] GÉANT. GÉANT Statetegy 2020 Implementing the Strategy, pp. 11-13. [Online]. Available from: http://www.geant.net/Resources/Media_Library/Documents/GEANT_Strategy2020-Implementation.pdf#search=Strategy2020-Implementation 2015.08.26

[9] J. Vojtěch , et al., "Device for Multicast of Optical Signals in the Internet and other Networks", European Patent EP2227911 (A2), 2014. [Online]. Available from: http://worldwide.espacenet.com/publicationDetails/biblio?CC=EP&NR=2227911A2&KC=A2&FT=D

[10] J. Vojtěch , et al., "Device for Multicast of Optical Signals in the Internet and other Networks", US Patent 8,582,967, 2013. [Online]. Available from: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=5&f=G&l=50&co1=AND&d=PTXT&s1=cesnet.ASNM.&OS=AN/cesnet&RS=AN/cesnet 2015.08.26

[11] M. Karásek , et al., "Modular set of devices for optical amplification of signal by raman fiber amplifier ", US Patent 8,630,035, 2014. [Online]. Available from: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=4&f=G&l=50&co1=AND&d=PTXT&s1=cesnet.ASNM.&OS=AN/cesnet&RS=AN/cesnet 2015.08.26

[12] J. Vojtěch , et al., "Modular kit of devices for variable distribution, mixing and monitoring of optical signals in the internet and other networks ", US Patent 8,948,590, 2015. [Online]. Available from: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-

bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=cesnet.ASNM.&OS=AN/cesnet&RS=AN/cesnet 2015.08.26

FPGA Based TCP Session Features Extraction Utilizing Off-Chip Memories

Satoshi Fuchigami

Graduate School of Information Science
Nagoya University
Nagoya,464-8601 Japan
Email: fuchigami@net.itc.nagoya-u.ac.jp

Hajime Shimada , Yukiko Yamaguchi

Information Technology Center
Nagoya University
Nagoya,464-8601 Japan
Email: {shimada, yamaguchi}@itc.nagoya-u.ac.jp

Hiroki Takakura
National Institute of Informatics
Tokyo,101-8430 Japan
Email: takakura@nii.ac.jp

***Abstract*—In recent years, unknown attacks, such as zero-day attacks and targeted attacks, have been increasing. These attacks are difficult to detect because the information gathered from already known attacks is not useful for their detection. An anomaly-based Network Intrusion Detection System(IDS) has the potential to find these attacks. However, almost all anomaly-based Network IDSs are implemented as software, so they cannot catch up with the growing network traffic. To alleviate this problem, there is Hardware/Software(HW/SW) cooperated Network IDS which migrates Transmission Control Protocol(TCP) feature extraction process to Field Programmable Gate Array(FPGA). However, the prior implementation is completed in FPGA, so that it cannot treat long TCP sessions because of shortage of memory blocks in FPGA-chip. In this paper, we propose TCP session feature extraction and cumulation by FPGA combining off-chip Ternary Content Addressable Memory(TCAM) and Dynamic Random Access Memory(DRAM) for HW/SW cooperated Network IDS. This approach uses these off-chip memories for buffering features while a TCP session continues. We present here the architecture design and implementation. We estimate that our system can manage 1,024K sessions simultaneously.**

Keywords–Anomaly Based Network IDS; FPGA; TCP Session Feature Extraction

I. INTRODUCTION

In recent years, cyber attacks have increased and more sophisticated, so that it is important to detect their invasion by monitoring network traffic. However, the amount of network traffic is growing rapidly and it requires more throughput to Network IDS. Furthermore, to alleviate these attacks, inspection of the internal network is also effective but it requires ten times larger throughput compared to Wide Area Network(WAN) gateway based inspection.

To resolve this problem, there are several studies using FPGA which is a re-programmable hardware for Network IDS. But past FPGA based Network IDS is only done in signature-based Network IDS. On the other hand, we have performed a study about HW(FPGA)/SW cooperated Network IDS [1] [2], which is implemented based on anomaly-based scheme and suited to detect increasing unknown attacks. We use FPGA for extracting TCP session features as a part of the system. It reduces the burden on the Network IDS software by migrating the feature extraction process to FPGA which occupies around 90% of CPU time [2]. However, in a previous implementation, the FPGA could not handle long and a large number of sessions because the implementation utilizes Random Access Memory(RAM) in the FPGA-chip whose capacity is quite small, i.e., 5.675Mbytes.

This paper describes a TCP session feature extraction system, which utilizes both off-chip TCAM and DRAM. The proposed system assists the Network IDS which uses PAYL [3] algorithm by implementing heavy feature extraction tasks into FPGA. When the system starts TCP session feature cumulation by a SYN packet, the proposed system prepares entry for buffering feature into both memories. Until TCP session finishes, the proposed system extracts features from TCP packets of the same session one by one and cumulates TCP session features using prepared entries. When the TCP session finishes, the proposed system outputs the TCP session feature to software side which is executed in general-purpose server machine.

The rest of this paper is organized as follows. Section II describes two tyeps of Network IDSs and a research about using FPGA for Network IDS. Section III addresses details of our proposal. Section IV explains the implementation. In section V, we estimate throughput for traffic feature extraction. Section VI concludes this research and suggest approaches for our future study.

II. RELATED WORK

There are two types of Network IDSs: signature-based Network IDS and anomaly-based Network IDS. The former detects attacks by comparing traffic data with signatures made from patterns of known attacks. This kind of method works well against known attacks but not against unknown attacks increasing today. Currently, these kinds of methods are widely used in the world and Snort [4] is one of the most famous software implementations.

The latter identifies attacks by statistically analyzing traffic features like clustering method such as K-means [5] and

One-Class Support Vector Machine(SVM) [6]. Those types of Network IDSs have the potential to detect unknown attacks, so that it is supposed to catch up to latest cyber attacks.

However, the current network traffic is already significant and continuously increasing. Network IDS is required to catch up with traffic in this environments. There are researches about implementing Network IDS using FPGA or Application Specific Integrated Circuit(ASIC) to alleviate this problem. Katashita et al. [7] proposed a 10Gbps throughput signature-based Network IDS using FPGA. This system inspects traffic data by traffic data signature matching method which is categorized into signature-based method. They also developed a tool which generates a circuit from Snort rules. On the other hand, hardware implementation of anomaly-based Network IDS is not generic because their detection algorithms are often difficult to implement in hardware.

III. Details of Our Proposed System

A. *Concepts of the System*

Future network traffic is expected to be subjected to many unknown attacks under a huge amount of traffic. To confront this situation, our proposing system aims to achieve high throughput and anomaly intrusion detection. As shown in Figure 1, in the existing anomaly intrusion detection method, traffic data are mirrored on switch and their copy are temporarily stored into storage. Then, intrusion detection process analyzes the stored data later. On the other hand, our proposed system aims at real time processing by reducing the burden of the server performing the analysis by extracting network traffic features on FPGA. This is a kind of HW/SW supported IDS. The FPGA also includes L2 switch functions, so that network traffic features extraction is done with port based distributed processing.

B. *Baseline and Functions*

We use Altera Stratix V GX (model: 5SGXEA7H2F35) FPGA. This board also has 20 Small Form factor Pluggable+(SFP+) ports and we already implemented L2 switch function to 8 ports of them with 1000BASE speed. The other SFP+ ports are unused to save hardware resources. The board also has TCAM/DRAM daughter board. The TCAM daughter board has 8 TCAM (model: IDT75K72100) chips and is configured as 144bit x 1,024K entries. The DRAM daughter board has 2 channeled DDR3-1600 SDRAM whose capacity is 16GB.

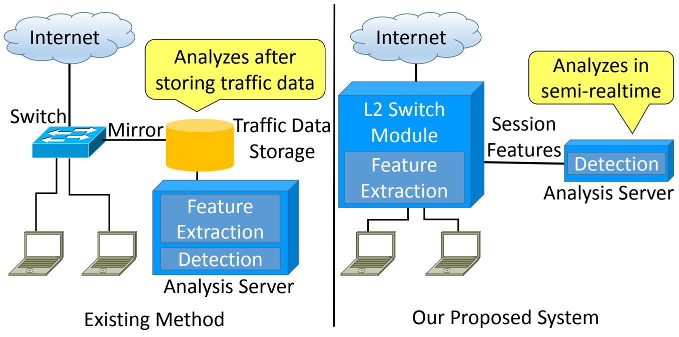


Figure 1. The difference between existing method and our method

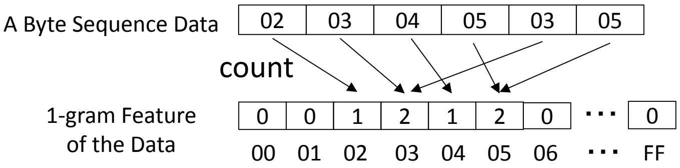


Figure 2. 1-gram Feature

We implemented feature extraction functionality [1]. When it receives a TCP packet, the Packet Feature Extraction Module extracts the header information and the 1-gram feature of the payload from the packet as shown in Figure 3. The header information contains the payload size and the src Internet Protocol(IP) address / src port number / dst IP address / dst port number, respectively. The 1-gram feature is the byte based value frequency of the payload as shown in Figure 2. Firstly, the payload is divided into 1-byte length and the counter for 1-gram feature counts appearance of 1 to 255 value in the playload. This feature used for PAYL detection algorithm, but it requires too many arithmetic resources.

Based on above packet based information extraction, our system cumulates them to create session features as the session continues. The features required to identify the session are shown as below.

- IP Address : Client / Server
- Port Number : Client / Server
- Total Packet Count : Each Communication Direction
- Total Payload Size : Each Communication Direction
- 1-gram Feature : Each Communication Direction
- Finish State : The State of Cumulation Termination

When cumulation of session features is finished, our system outputs them. There are several patterns to terminate the cumulation, so that we prepare the Finish State. It indicates three patterns of termination that are termination by FIN Flags, termination by RST Flags, and termination by lack of buffer capacity.

C. *Data Structure for Session Feature Extraction*

We utilize off-chip memory to record session features. One TCAM and DRAM entry of fixed size storage area are assigned for each session. When the session starts, these entries are assigned for cumulation. Until the session finishes, features are cumulated using the entries as a buffer of partial cumulation. After the session finishes, assigned entries are deleted by sending their content to the analysis server. The contents of TCAM and DRAM entries are as follows.

TCAM Entry
    IPaddrLow(4bytes) , IPaddrLowPort(2bytes) ,
    IPaddrHigh(4bytes) , IPaddrHighPort(2bytes) ,
    IsIPaddrHighServer(1byte) ,
    DRAMaddr(4bytes)
DRAM Entry
    1-gramPayloadFeature (1,024bytes × 2)
    PacketCnt (4bytes × 2)
    PayloadSize (4bytes × 2)
    SessionState (1byte)

TCAM entry works as an index of DRAM entry. A search key of TCAM entry consists of IPaddrLow, IPaddrLowPort,

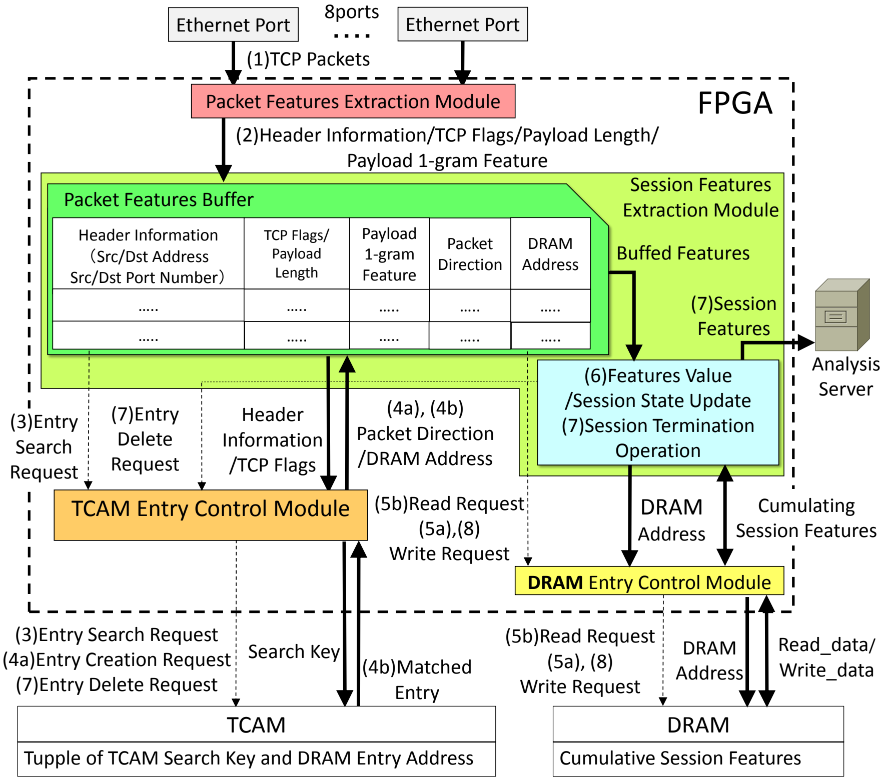


Figure 3. The behavior of Session Features Cumulation

IPaddrHigh, and IPaddrHighPort. IPaddrLow is the smaller one of either client IP address or server IP address in 32-bit numeric order. To save number of entries between bidirectional communication, we sorted IP addresses with 32-bit numeric order for index. By sorting this way, packets of both directions of a same session are assigned to the same entries. IPaddrLowPort is the port number of the IPaddrLow host and IPaddrHighPort is the port number of IPaddrHigh host. IsIPaddrHighServer is the identifier of the server which is required to identify the server after IP address sorting. With these 5 fields, we can identify the session. The DRAMaddr is the address of the DRAM entry which keeps detailed cumulating session features like 1-gram features of the session. In this way, we combine TCAM and DRAM for the session identification and the session features cumulation buffer. This organization can reduce consumption of TCAM and DRAM entries.

DRAM entry stores cumulating features while the session is in progress. The PacketCnt is the cumulation of packet count and the PayloadSize is the cumulation of payload size. The 1-gramPayloadFeature is the cumulation of 1-gram feature of all packets. These three fields are separated by communication direction.

D. The Operation of the System

Figure 4 shows a flowchart of the session feature cumulation process when a TCP packet arrives. Figure 3 shows a block diagram of the implementation which executes the session feature cumulation process. In our previous study, we implemented Packet Features Extraction Module as shown in Figure 3. In this study, we modified Session Feature Extraction Module and developed TCAM Entry Control Module, and DRAM Entry Control Module.

The operation of the system is as follows.

(1) When an Ethernet Port receives a TCP packet, the packet data comes into the system. Then, the Packet Features Extraction Module extracts header information, TCP flags, payload length, and 1-gram feature of the payload from the packet.

(2) It requires some latency to access the TCAM and the DRAM because of their access latency. Therefore, the packet features are buffered in Session Feature Extraction Module.

(3) The header information and TCP flags are sent to the TCAM Entry Control Module. Then, the module calculates the search key by comparing two IP addresses of the header information as unsigned 32-bit integer order and defines IsIPaddrHighServer by recording the information whether IP addresses are sorted or not. If the TCAM port is available, the module sends the search request to the TCAM.

(4) (4a) If the matched entry does not exist in the TCAM and the packet is a SYN packet, we treat it as a start of the session and the module generates new entry for the new session. The module calculates the address of the new DRAM entry and initializes the contents of the TCAM entry. Also, it generates the packet direction information. Both the DRAM address and the packet direction information are sent

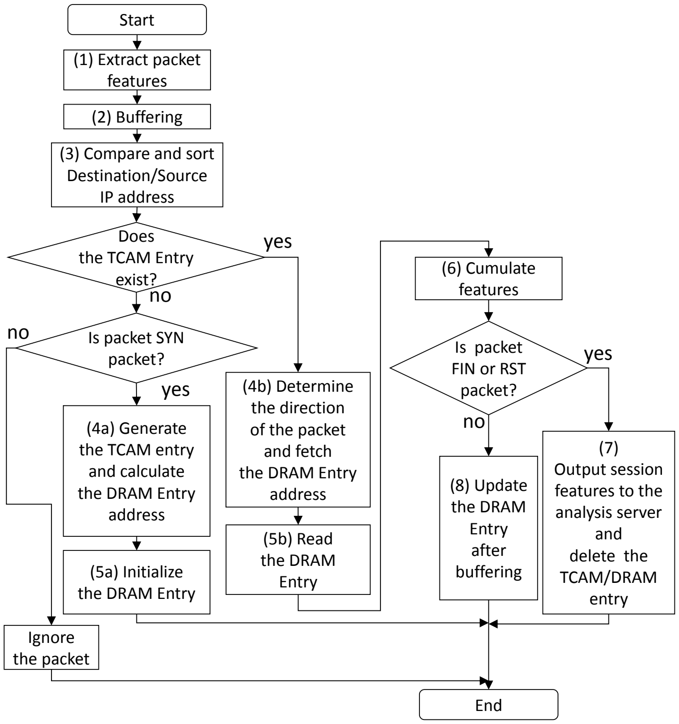

Start
(1) Extract packet features
(2) Buffering
(3) Compare and sort Destination/Source IP address
Does the TCAM Entry exist?
yes
no
Is packet SYN packet?
no
yes
(4a) Generate the TCAM entry and calculate the DRAM Entry address
(5a) Initialize the DRAM Entry
(4b) Determine the direction of the packet and fetch the DRAM Entry address
(5b) Read the DRAM Entry
(6) Cumulate features
Is packet FIN or RST packet?
yes
no
(7) Output session features to the analysis server and delete the TCAM/DRAM entry
(8) Update the DRAM Entry after buffering
Ignore the packet
End

Figure 4. Operation Flowchart

to Session Features Extraction Module and recorded in the Packet Features Buffer.

(4b) If the matched entry exists, there is a session which is already started. The packet direction is determined by sorting and the IsIPaddrHighServer of the matched entry. Both the DRAM address and the packet direction are recorded in Packet Features Buffer.

(5) (5a) (comes from (4a)) If the session is the new session, the new session feature is written into the DRAM as a new entry through the DRAM Entry Control Module.

(5b) (comes from (4b)) Session Features Extraction Module sends a read request to the DRAM through the DRAM Entry Control Module, to read partially cumulated session features.

(6) After reading out the DRAM entry, the module cumulates the 1-gram payload feature to the 1-gramPayloadFeature, cumulates the payload length to the PayloadSize, and increments the PacketCnt. If the packet is a FIN or a RST packet, the following operation becomes (7). Otherwise, the following operation becomes (8).

(7) If the packet is a FIN or a RST packet, the session is finished with this packet. The cumulated session features are sent to the analysis server. Also the TCAM entry delete request is sent to the TCAM through TCAM Entry Control Module. This operation is also activated when the system consumes the entire TCAM and the DRAM entries, and it has to terminate the oldest cumulating session.

(8) The module updates the fields of the DRAM entry with newly cumulated values.

IV. Implementation Results

In this study, we implemented the following three modules, Session Features Extraction Module, TCAM Entry Control Module, and DRAM Entry Control Module by the Verilog Hardware description language(HDL). However, we have not finished the implementation of the function of entry deletion nor termination of oldest cumulating session when the system consumes all TCAM and DRAM entries.

We synthesized those modules with Altera Quartus II 13.1. Table I and II show the result of the current FPGA resource utilization. According to Table 1, the modules in this implementation occupy 42% of whole logic elements because we implemented 256 adders for cumulating 256 1-gram payload feature simultaneously. Therefore, logic elements of whole system are 96% of all logic elements, so that there is no space to optimize for operating clock frequency and no room for additional functions. Hence, we do not include operating clock frequency results. Also, there is still unimplemented functionality, so that we have to improve our current implementation to reduce the usage of logic elements. According to Table II, the three modules occupy 22% of all registers because we implemented Packet Feature Buffer as a register array. If we increase the number of buffers to catch up with higher throughput (e.g., 10GBASE × 8), we have to re-implement it with block RAM. Furthermore, the current implementation accesses the DRAM when it receives a TCP packet. This becomes a possible bottleneck of the system, so that we are just considering some type of cache.

TABLE I. LOGIC ELEMENTS UTILIZATION OF COMPONENTS

Components	used	total	usage
Whole System	225,474	234,720	96%
The Three Modules	97,902	234,720	42%

TABLE II. REGISTERS UTILIZATION OF COMPONENTS

Components	used	total	usage
Whole System	438,482	938,880	47%
The Three Modules	204,909	938,880	22%

There are still many difficulties, but our system can currently handle a total of 1,024K entries of buffers. Therefore, it can handle 1,024K sessions simultaneously.

V. Throughput Estimation

We estimated the throughput of current implementation. In the worst case, the proposed system accesses the TCAM twice for searching and making a new entry. After that, it accesses the DRAM twice for reading and writing entries. In the proposed system, we made pipeline stages to enable accessing the TCAM and the DRAM in parallel, so that we only have to consider whether either of them is a bottleneck or not. Firstly, we estimate throughput from DRAM side because DRAM becomes bottleneck in many systems. DRAM access time for transmitting given data size (byte) is shown as

$$tRAS + tRCD + tCAS + tCLK \times \frac{data\_size}{8} \qquad (1)$$

Where tRAS is row access strobe time, tRCD is row to column delay time, tCAS is column access strobe time, and tCLK is clock cycle time of data transfer. By substituting typical values of DDR3-1600 SDRAM and the data size for one session, the above formula is translated as follows.

$$45ns + 12.5ns + 12.5ns + 1.25ns \times \frac{2065}{8} = 393ns \quad (2)$$

Note that the above value is DRAM access time for one access. The proposed system requires two DRAM accesses in one packet processing, so that the substantial DRAM access time becomes twice that value. But our system has two DRAM channels, so that if we adequately interleave DRAM access, the DRAM throughput becomes twice that much value. Thus, DRAM access time per one packet processing becomes 393ns in our system. The DRAM access time is 393ns and packet throughput is 2.54Mpps (packet per second). The data throughput is related to the size of a packet and number of DRAM accesses per packet as follows.

$$\frac{packet\_throughput \times average\_packet\_size}{access\_count\_for\_one\_procedure} \quad (3)$$

So, if we assume 1500 bytes packets, the throughput becomes 38.1Gbps. If we assume 64 bytes short packets, the throughput becomes 1.62Gbps. Thus, we have to consider some filtering scheme for huge amount of short packets. On the other hand, the TCAM which we utilized can treat 250M search per second and it is much larger than that of the DRAM. So, the TCAM does not affect throughput in the proposed system.

VI. Conclusions and Future Work

We proposed the method of TCP session features extraction for anomaly-based Network IDS by FPGA using off-chip memories. We implemented the proposal to FPGA and confirmed that we can implement 1,024K session treatable system. But we also confirmed that the current implementation consumes almost all FPGA resources, so it requires further updating to implement additional functions and raise throughput.

In the future, we will reduce hardware resource consumption of implementation by modifying feature cumulation circuit to calculate in multi cycles. This alteration will enable us to implement additional functionality. Moreover, we will also improve the algorithm of the feature extraction processes to raise throughput of the system. Finally, we will evaluate the real throughput of the system by operating it in real traffic environment.

Acknowledgment

This research is aided by R&D of detective and analytical technology against advanced cyberattack, administered by the Ministry of Internal Affairs and Communications.

References

[1] S. Yanase, H. Shimada, Y. Yamaguchi, and H. Takakura, “Network access control by FPGA-based network switch using HW/SW cooperated IDS,” TECHNICAL REPORT OF IEICE, vol. 114, no. 286, 2014, pp. 91–96.

[2] S. Yanase, H. Shimada, Y. Yamaguchi, and H. Takakura, “Implementation of FPGA section for anomaly detection acceleration by HW/SW cooperation (in Japanese),” TECHNICAL REPORT OF IEICE, vol. 114, no. 116, 2014, pp. 75–80.

[3] K. Wang and S. J. Stolfo, “Anomalous payload-based network intrusion detection,” in Recent Advances in Intrusion Detection. Springer, 2004, pp. 203–222.

[4] Snort, “Snort.Org,” http://www.snort.org, Accessed: 2015-8.

[5] G. Münz, S. Li, and G. Carle, “Traffic anomaly detection using k-means clustering,” in GI/ITG Workshop MMBnet, 2007.

[6] R. Perdisci, G. Gu, and W. Lee, “Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems,” in Data Mining, 2006. ICDM’06. Sixth International Conference on. IEEE, 2006, pp. 488–498.

[7] T. Katashita, Y. Yamaguchi, A. Maeda, and T. Kenji, “FPGA-based intrusion detection system for 10 gigabit ethernet,” IEICE transactions on information and systems, vol. 90, no. 12, 2007, pp. 1923–1931.

Reliable Assurance Protocols for Information Systems

Mahalingam Ramkumar

Computer Science and Engineering
Mississippi State University
Mississippi State, MS
Email: ramkumar@cse.msstate.edu

Somya D. Mohanty

Social Sciences Research Center
Mississippi State University
Mississippi State, MS
Email: somya.mohanty@ssrc.msstate.edu

***Abstract*—The assurances provided by an *assurance protocol* for any information system (IS), extend only as much as the integrity of the assurance protocol itself. The integrity of the assurance protocol is negatively influenced by a) the complexity of the assurance protocol, and b) the complexity of the platform on which the assurance protocol is executed. This paper outlines a holistic Mirror Network (MN) framework for assuring information systems that seeks to minimize both complexities. The MN framework is illustrated using a generic cloud file storage system as an example IS.**

Keywords: ***Clark-Wilson Model, System Integrity, Ordered Merkle Tree, Cloud Storage***

I. Introduction

Information systems (IS) are composed of a variety of hardware and software components that create, exchange, process, and dispose data. From a broad perspective, assuring the operation of an IS is a process involving *verification of self-consistency* of all critical internal states of the IS. From this perspective, the assurance mechanism (or the *assurance protocol*) itself can be seen as software that

1) verifies self-consistency of IS data, and
2) reports consistency/inconsistency to entities that interact with the IS.

For example, some of the simple self-consistency checks that will need to be performed by an assurance software for an accounting system include a) that available balance in an account is incremented by the amount deposited, or decremented by the amount withdrawn; b) that transfer of an amount a from an account A to account B results in increase in account B balance by a, and reduction in account A balance by $a + x$ (where x is a service charge), etc.

The assurances offered by the assurance software are, at best, only as good as the integrity of the assurance software itself. In general, the higher the complexity of any hardware/software component, the higher the possibility of presence of undesired (malicious or accidental) functionality [1]. Consequently, it is important to minimize both a) the complexity of the assurance software, and b) the complexity of the platform in which the assurance software is executed.

The motivation for the proposed approach to assure ISes stems from the fact that the assurance software for an IS can be *substantially simpler* than the IS software; consequently, the assurance software can be easily executed on a dedicated *high-integrity-low-complexity platform* (Figure 1), that is *completely isolated* from the actual IS. In the proposed approach, the platform for execution of the assurance software is constrained to be a homogeneous network, composed of a single type of *low complexity building block*. The mirror network (MN) model outlined in this paper involves assembling any number of such building blocks — hereinafter referred to as *MN modules* $\mathbf{T}$, into a network that a) mirrors critical IS states; and b) executes IS-specific assurance protocols for checking/reporting self-consistency of IS states.

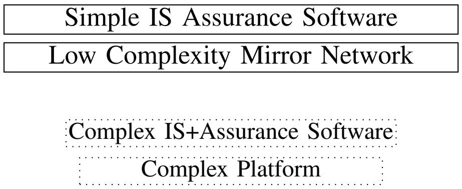

Simple IS Assurance Software
Low Complexity Mirror Network
Complex IS+Assurance Software
Complex Platform

Figure 1. MN Model (top) vs Conventional Model (bottom)

The main contributions of this paper are a) an overview of simple *generic* functional components of MN building blocks, that permit them to

1) assemble themselves into an MN, and
2) jointly execute the assurance protocol

for *any* IS, and b) illustration of the process of assurance software design, using a generic cloud file storage service as an example.

The rest of this paper is organized as follows. Section II provides a broad overview of the MN model and its relationship to the Clark-Wilson (CW) system integrity model. Specifically, while the CW model is applied directly to the IS to be secured, the MN model can be seen as a variant of the CW model, *applied to the assurance protocol* for the IS. This feature has has two important advantages. Firstly, the IS assurance software for an IS can be substantially simpler than the IS. Secondly, the assurance software for different ISes tend to be more similar than the ISes themselves — making it possible to reuse a small number of simple functional components to realize assurance software for different ISes.

Application of the MN model to an IS results in a simple *MN specification*, which *is* the assurance software for the IS, intended to be executed on a special platform — a mirror network. Section II-B outlines the mechanism for MN deployment. Section II-C reviews some simple built-in functional components in MN modules which can be leveraged to deploy MNs, and permit them to jointly execute the assurance software. Section III describes various steps involved in designing the MN specification (the assurance software) for an example IS — a cloud file storage system. Conclusions and a brief comparison between the MN approach and an alternate approach in the literature [2] (also based on a specification of low complexity modules) are offered in Section IV.

II. MN Model for Information Systems

The Clark-Wilson (CW) [3] model for system integrity is characterized by constrained data items (CDI), unconstrained data items (UDI), transformation procedures (TP), integrity verification procedures (IVP), and CW-tuples, where

1) **CDIs** are unambiguously labeled system states whose integrity needs to be assured;
2) **IVPs** determine if the current state of all CDIs represent a valid IS state;
3) Only "well-formed" **TPs** can *modify* or *create* CDIs;
4) **UDIs** can not be constrained as they are external inputs to the system; they are the primary triggers for creation / modification of CDIs.
5) **CW tuples** of the form (user, TP, CDIs/UDIs) specify which user process is allowed to execute which TP, and which CDIs are affected by the TP.

A TP is well-formed if it is guaranteed to always take the system from one correct state to another correct state. If at any time, the correctness of the IS state is demonstrated by an IVP, and thereafter, if only well-formed TPs are used to modify/create CDIs, it follows from induction, that the system is guaranteed to always remain in a correct state. In the CW model, the correctness of IVPs, TPs and CW-tuples are assumed to be certified by a "security officer."

A. MN Model

In the (ρ, μ, ν) MN model for an IS S with ρ CDI database types, μ message types and ν event types

1) One-way functions of crucial IS S states are grouped together into **CDI databases** of ρ different types.
2) *Events* (of ν types) trigger a) modifications to the CDI databases, and/or b) creation of **MN messages** (any of μ types).
3) Events can be external or internal. External events are UDIs. Internal events are triggered by **MN messages**, created by an external or internal event.
4) Each event type is associated with a **TP**, specifying a list of pre-conditions (for execution of the TP) and post-conditions (following execution of the TP).

To summarize, the MN model for an IS S is simply a specification of ρ types of CDI databases, ν event-types/TPs and μ types of MN messages. The designer of the MN has complete freedom in choosing convenient ρ, μ and ν, depending on the nature of the IS (in the example MN for a cloud file storage system described in a later section we choose $\rho = 3, \mu = 4$ and $\nu = 17$). The MN specification for an IS S is represented as a *static MN-rules database*, which is **the static assurance "software"** for IS S. More specifically, "execution of the assurance software" is simply execution of the ν TPs specified in the MN-rules database. A static cryptographic commitment to the MN-rules database, say S', doubles as the *identity* of the MN (the platform) deployed to execute the assurance software for the IS S.

B. Execution of Assurance Software

The MN S' (deployed for assuring IS S) is a dynamic network, composed of any number of MN modules (which become *members* of the MN). For an MN with ρ different CDI database types, members with ρ different *roles* will exist. The total number of members (MN modules) $d(t) = \sum_{i=1}^{\rho} n_i$ (or n_i members with role i) is dynamic (need *not* be specified *apriori* in the MN rules database). All MN modules possess identical functionality; the differences between modules are merely their unique identities and secrets. Within the context of MN S', each module is assigned a unique role based member identity, depending on the CDI database type maintained by the member.

Apart from the $d = \sum_{i=1}^{\rho} n_i$ members (that track CDI databases), every MN includes a special module regarded as the *creator* of the MN. The MN creator is responsible for inducting other modules into the MN as members. Unlike the $d = \sum_{i=1}^{\rho} n_i$ modules that track dynamic CDI databases, the MN creator module tracks a dynamic MN membership database. For example, if MN modules with identities $\Pi_1 \cdots \Pi_d$ have been inducted into the MN S', and assigned role based identities $(X_1 \cdots X_d)$ respectively, the membership database maintained by the MN creator module will have d records of the form (X_i, Π_i). The role-based member identities like X_i explicitly indicate (using reserved bits) the role of the member.

From the perspective of a MN member X with role i, "tracking" a CDI database of type i involves a) unambiguously identifying the TP to be executed in response to an event, b) verifying pre-conditions, and c) imposing post-conditions. Preconditions can be i) existence/nonexistence of specific records in X's CDI database; and/or ii) receipt of a MN message. Post-conditions can be i) updates to specific records in its CDI database; and/or ii) creation of an MN message.

During regular operation of the IS S, external events (UDIs) are conveyed to the MN. This is the *only* link between the IS S and the MN S'. A member X in the MN, triggered by an event, executes a TP, which can result in modification to one or more CDI database records of X, and/or creation of a MN message from member X to another member Y. The MN message so created, triggers execution of a TP by Y, which can trigger modification to CDI records of Y and/or creation of a message addressed to a MN member Z, and so on.

C. Generic MN Module Functions

To reduce the complexity of the platform (the MN), MN modules are *deliberately* constrained to be able to perform only simple sequences of logical and hash operations that demand only modest and constant memory size for execution. Fortunately, the versatility of cryptographic hash functions renders them more than adequate for

1) realizing simple security protocols for tracking the integrity of dynamic databases, and

2) facilitating authentication and privacy of a) MN messages between MN modules, b) UDIs from external entities to MN modules; and c) state reports from MN modules to external entities.

In other words, simple generic (IS-independent) protocols built-in into MN modules provide the foundation for richer protocols necessary for deployment of MNs, *and* execution of IS-specific TPs.

1) Index Ordered Merkle Tree: From the perspective of MN modules, any database is seen as a collection of (index,value) tuples (or records). Protocols for maintaining an index ordered Merkle tree (IOMT) [4], [5]-[7] permit resource limited modules that store only a single hash — the *root* of the IOMT — to perform reliable database operations for reading/ updating/ inserting/ deleting uniquely indexed records/tuples in a *virtually* stored database. In other words, the actual database of records can be stored in any convenient (and possibly untrusted) location – for example, by the untrusted IS. For a virtually stored database with N records, each basic database operation will only require $\mathbb{O}(\log_2 N)$ hash evaluations by the module (for example, 40 hashes for a database with a trillion records). IOMTs can also be used to represent nested tuples — where the value v in tuple (a, v) can itself be the root of an IOMT.

In databases represented using an IOMT, a record of the form $(idx, val = 0)$ is a *place-holder*, indicating "absence of information" regarding index idx. The main difference between an IOMT and the better known "plain" Merkle hash tree [8] is that the IOMT includes protocols for *insertion/deletion* of place holders to guarantee uniqueness of indexes. Protocols for updating/reading records using an IOMT are, however, identical to that of a "plain" erkle tree.

Simple built-in capability to execute IOMT protocols confer MN modules (which store only a single hash) with the ability to a) verify pre-conditions like existence of a record (f, v), non-existence of a record for index f (or equivalently, existence of place-holder $(f, 0)$), in the virtually stored CDI database and b) modify the IOMT root stored inside in accordance with modifications made to one or more virtually stored CDI tuples, as demanded by post-conditions of a TP.

Specifically, MN modules use their ability to execute IOMT protocols to reliably perform

1) read/write/insert/delete operations in dynamic CDI databases for purposes of verifying pre-conditions and imposing post-conditions; each MN member (module) tracks one CDI database;
2) read/write/insert/delete operations in the dynamic MN-membership databases for inducting/ejecting modules into/from the MN; there is only one such database for each MN, maintained by the MN creator module;
3) read operations on the static MN-rules databases; the same database is referred to by every member of the MN. As all members of the MN S' are initialized with the same value S', they will only honor TPs in this common database.

2) Authentication and Privacy: Several key distribution schemes [4], [9] – [11] for establishment of pairwise secrets have been explicitly designed for scenarios involving severely resource limited participants. For example, the MLS protocol [11] will require every module to store only a single secret, and evaluate a single hash to compute a pairwise secret with any other entity. Two modules X and Y with secrets K_X and K_Y respectively can compute a common secret $h(K_X, Y)$ or $h(K_Y, X)$ depending on which entity has access to a pair-wise *public* value

$$P_{XY} = h(K_X, Y) \oplus h(K_Y, X) \tag{1}$$

If X has access to the public value the pair-wise secret is computed by X as $h(K_X, Y) \oplus P_{XY} = h(K_Y, X)$, which can be computed by Y by hashing its secret. The number of pair-wise public values required is not a serious concern as they can be stored virtually (outside the module).

Pair-wise secrets facilitated by schemes like MLS can be used for computing hashed message authentication codes (HMAC) for a) mutual authentication, and b) protecting privacy of secret components in messages. In the MN model, the built-in ability of MN modules to compute pairwise secrets are leveraged for the following specific purposes:

1) mutual authentication of message exchanges between MN modules (potential members and the MN creator) to join an MN;
2) mutual authentication of MN messages between MN-members,
3) mutual authentication and privacy of communications between MN members and external entities (who convey UDIs, and may query a MN member for the state of the MN)

A MN message can also be a *self-message* — from a member to itself. Self messages from a member X are authenticated using a self-secret S_X known only to X (randomly generated by X). Self secrets can also be used by a MN module to encrypt *other* secrets entrusted to the module (and store encrypted secrets virtually).

For example, external entities can employ the MN for distributing secrets. Specifically, let an external entity u share a secret K_{xu} with a MN member X. Entity u can utilize the following simple protocol to share a secret K with any number of entities, specified indirectly through a *context* f. The entity u sends values c, s_u, f related as

$$c = h(K, f) \text{ and } s_u = h(K_{xu}, c) \oplus K \tag{2}$$

The value c is a commitment to both the secret K and the context f, and serves as a public identifier for the secret K; s_u is the link-encrypted version of the secret K. The module (which can readily compute K_{xu}) computes $K = h(K_{xu}, c) \oplus s_u$, verifies that $c = h(K, f)$, and uses its self-secret K_x, to re-encrypt the secret K for storage as $s = h(K_x, c) \oplus K$ (more specifically, a tuple (c, s) is added to the CDI database tracked by the module).

An entity w with whom the module shares a secret K_{xw} may receive the secret K under some conditions. Firstly, the module X should be a) convinced of the existence of the record (c, s), and b) provided a value f satisfying

$$c = h(h(K_x, c) \oplus s, f). \tag{3}$$

In addition, if a MN-specific rule relates w and the context f (for example, a rule can be "existence of a record for index w in an IOMT with root f.") the module may output values c and $s_w = h(K_{xw}, c) \oplus K$ to entity w. Entity w (who has access to secret K_{xw}) can decrypt the secret and check its integrity by verifying that $c = h(K, f)$.

Given that simple protocols to support such generic functions can be easily implemented even in severely resource limited modules, in the rest of this paper, we focus on the process for designing the MN rules database (the "assurance software" for the IS) with an illustrative example.

III. MIRROR NETWORK DESIGN EXAMPLE

The creation of the MN rules database for an IS S can be seen as a process consisting of the several steps like 1) identification of desired assurances; 2) identification of the subset of data items (or one-way functions of data items) that need to be constrained in order to realize the desired assurances; 3) choice of ρ types of CDI databases, each possibly with a different interpretation of the index and value fields. 4) enumeration of ν event types and μ message types; and 5) specification of TP for each event in the form of pre/post-conditions. All components of the MN specification become leaves of a static IOMT with root S'.

A. Desired Assurances for a Cloud Storage Service

Cloud storage services offer a convenient way for users to share files between multiple platforms – even platforms owned by different users. From a security perspective, users of such a service desire assurances regarding the *integrity*, *privacy*, and *availability* of files.

In such a system, software running on end-user platform may periodically upload new files, or newer versions of existing files to the service. Users may also be able to specify access control lists (ACL) for every file they own, indicating read/write access restrictions for other users. For ensuring privacy of files, the files may be encrypted by users. As users who share an encrypted file will need to share the file-encryption secret, and as users may not have an out-of-network strategy for exchanging such secrets, the file storage service itself should cater for secure mechanisms for conveying file encryption secrets to authorized users.

The desired assurances for a remote file storage service [6] can be summarized as follows:

A1 The service will not alter files; only users explicitly granted the permission (by the owner) to modify the file can do so.
A2 File encryption secrets will not be abused by the service.
A3 The service will not modify ACLs, and strictly abide by ACL permissions.
A4 Only the latest version of the file will be provided by the service to authorized users (except when explicitly queried for an earlier version).
A5 After an ACL has been modified by an authorized user, the older ACL should not be used to determine the access privileges.
A6 A user with legitimate access rights will not be improperly denied access to the file.

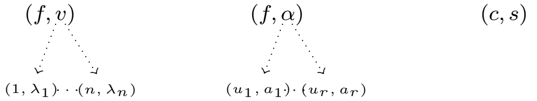

(f, v)
$(1, \lambda_1) \cdots (n, \lambda_n)$
(f, α)
$(u_1, a_1) \cdot (u_r, a_r)$
(c, s)

Figure 2. Structure of records in CDI databases. File database (left), ACL database (middle) and encryption-secret database (right).

B. Constrained Data Items and MN Roles

The problem of assuring the integrity of a file can be reduced to that of assuring the integrity of the cryptographic hash of the file. Likewise, assuring the privacy of contents of a file can be reduced to assuring the privacy of a file-encryption secret, and that it is made available only to authorized users included in the ACL for the file. Thus, from the perspective of realizing the desired assurances, the CDIs for the MN are 1) file hashes corresponding to every version of every file; 3) the ACL for every file; and 3) all file encryption secrets. The CDIs can be seen as three types of databases, with possibly different interpretations of the index field and value field in the database records.

1) File databases, indexed by unique file indexes;
2) ACL databases, also indexed by file indexes;
3) File-encryption-secret database where the index is a "key" identifier c.

In a record (f, v) in the file database (Figure 2, left), f is a unique file index, and v is the root of a nested IOMT. A record (n, λ_n) in the nested IOMT provides information λ_n regarding the n^{th}version of the file index f. The ACL database (Figure 2, middle) has records of the form (f, α) where α is the root of an IOMT capturing the ACL for file f. A record (u_i, a_i) in the nested IOMT with root α indicates that user u has access permission a (for file index f). For example, $a_i = 1$ for read access, $a_i = 2$ for read-write access and $a_i = 3$ for write-access to the ACL (users with access level 3 can even change the ACL α for f). In a record (c, s) (Figure 2, right) in the file-encryption-secret database the value s is an encrypted version of a file encryption secret K_f. Specifically, the secret is encrypted using the self secret of the module tracking the CDI database. The index computed as $c = h(K_f, f)$ is simultaneously a commitment to both the secret K_f and the *context* f. The implication of the context f is that secret K_f should be made privy only to users with access level 1 or higher in the ACL for file f.

Corresponding to the three different CDI database types, the MN employs members with $\rho = 3$ different member roles, say role F (for file version databases), role S (file encryption secrets) and role A (ACL). Any number of members may exist for each role, each maintaining data pertaining to different non overlapping ranges of file indexes.

C. UDIs

Modifications to the CDI databases are triggered by UDIs emanating from users of the service. Specifically, corresponding to creation of new files, new file indexes are added to the CDI databases. Corresponding to updates to a file with index f a new version record is added. File owners may also submit

ACLs for files (CDI α, the ACL root), delete files (remove record for file f), submit file encryption secrets, etc.

External entities (software running on end-user platforms) interact with S-role MN members to i) convey hashes and/or secrets corresponding to new file versions, changes to ACLs, and ii) query the MN for file hashes and secrets. The results of the query can then be compared with files provided by the service.

As in the CW model, requests from users, which are UDIs (as any user can send any request — even unauthorized ones), should be logged. As users have to share a secret with members with role S (as such members convey/accept encrypted file-encryption secrets to/from users), it is convenient to let members with role S to also maintain a log database. In this paper, in the interest of keeping the discussions simple, we shall ignore the log database.

D. MN Messages

Four types of MN messages can be defined: types ACL (to report ACL privilege), AU (to update ACL), VU (to add a new version), and FP (to report file parameters).

A message $ACL_{X,Y}(f, a, u)$ (from X to Y) of type ACL, can be created by a member X (with role A) and delivered to a member Y (with role F or S), indicating that user u has access permission a for file f. To generate such a message, X merely needs to confirm the existence of a record (f, α) in its CDI database, and the existence of record (u, a) in an IOMT with root α. An ACL message for a file f with $u = a = 0$ can be created only if the ACL IOMT root α for f is 0 (as we shall see later, such a message is used to trigger removal of (all versions of) file f).

A message $AU_{X,Y}(f, u, \alpha)$, can be created by S members and sent to A members, indicating a request from a user u to update the ACL for file f. While any user u can request an S member to create such a message, the AU message will be honored by the A-role member only if user u has access right $a = 3$ for file f. Such a user can also set the value α to zero (to request deletion of the file f).

A message $VU_{X,Y}(f, u, \lambda)$, represents a request to create a new version of file f. This can be created by S members and sent to A members to check if user u has the necessary access permission. After ensuring sufficient access rights, A members can then create another $VU()$ message addressed to an F member. Once again, while any user can trigger the S member to create a VU message the VU message will be honored by the A-member only if u has access right 2 or higher for file f. VU messages created by A members (in response to a VU message from a S member) trigger F members to appropriately modify their CDI database record for file f.

Message of type $FP_{X,Y}(\cdots)$ conveying parameters for version q of file f are created by F members and conveyed to S members who may then relay the contents of the message to users of the service.

E. Events and TPs

The MN rules database specifying pre-conditions and post-conditions for all TPs (corresponding to each of the 17 event types 01 to 17) is depicted in Table I. Each event type is associated with role type(s) of member(s) who may respond to the event (role type indicated in parentheses alongside the event number in column 1). Column 2 lists UDIs (as $\{\cdots\}$), and other inputs (OI) necessary to execute the TP. Only TPs corresponding external events accept UDIs. TPs for internal events are triggered by MN messages.

Columns 3 and 4 depict the pre-conditions and post-conditions, respectively. A MN message in pre-conditions (column 3) indicates receipt of the message. A message in post-conditions (column 4) implies the need to *create* such a message. As the events are listed from the perspective of a member X, all messages in pre-conditions indicate only the sender of the message (the receiver is always X); all messages in post-conditions indicate only the receiver (the sender is always X). As can be seen from the table, events 01, 02, 06, 11, and 12 are external events as they are *not* triggered by MN messages.

A tuple (x, y) in pre-conditions indicates the presence of record (x, y). $(x, 0)$ represents absence of record for x (or presence of place-holder for x). $(x, y) \rightarrow (x, y')$ in post-conditions implies the need to update the IOMT root to account for the update to the value of record index x (from y to y'). $(x, (y, a))$ in pre-conditions indicates the presence of a nested record (y, a) for a record with index x. $(x, (y, a) \rightarrow (y, a'))$ in post-conditions indicates the need to update the nested record (and accordingly, update the IOMT root). $s \rightsquigarrow s'$ indicates that s' and s are related through symmetric encryption.

A user reserves a file index f by creating event 01, which generates a AU message, which becomes input to event 08, which outputs a AU message, that becomes input to event 04, which results in a confirmation message to the user. A user u can also trigger event 01 to modify the ACL for file f. In this case, the AU message from event 01 triggers event 09. Only if the user has access level 3, the ACL is updated, and a AU response is created, which triggers event 04, to send a message to the user, confirming successful ACL modification. If user u does not have sufficient access right, event 11 is triggered to create a ACL message, which triggers event 03, which creates a message informing the user of his/her access right. If the file does not exist, event 12 is invoked instead to create the ACL message. If the user does not have access, or if the file does not exist, the user receives a message conveying values $\{f, u, 0\}$. Thus, if the user does not have access to a file f, the user does not even get to know if file f exists. A user can request deletion of a file by updating the ACL to 0. Following this, event 12 can be invoked with $u = 0$ to create a ACL message, that triggers event 13, to delete all versions of file f.

A user u can convey a new file version, by invoking event 02. The VU message invokes event 10. Only if the user has write access, is the output VU message created. If the update is the first version of the file, the VU message triggers event 14. Else, it triggers event 15. Both output a FP message which triggers event 05, resulting in a acknowledgement to the user, that the update was successful. If the user did not have access, or has read-only access, as earlier, event 11 or 12 can be triggered to convey this fact to the user.

Any user can provide a secret to the MN by triggering

TABLE I. MN Rules for Cloud Storage Service MN with $\rho = 3$ types of member roles, $\nu = 17$ types of events (and TPs), and $\mu = 4$ types of MN messages. Pre/post-conditions for 17 events are listed for a member with identity X.

Events	UDI / OI	Preconditions	Post-conditions
$01(S)$	$\{f, \alpha\}, Y, u$		$AU_Y(f, u, \alpha)$
$02(S)$	$\{f, \lambda\}, Y, u$	Y type Λ	$VU_Y(f, u, \lambda)$
$03(S)$		$ACL_Y(f, u, a)$	$\{f, u, a\}_X$
$04(S)$		$AU_Y(f, u, \alpha)$	$\{f, u, \alpha\}_X$
$05(S)$		$FP_Y(f, q, \lambda, q')$	$\{f, q, \lambda, q'\}_X$
$06(S)$	$\{f, c, s'\}$	$s' \rightsquigarrow K, c = h(K, f) \neq 0$	$K \rightsquigarrow s, (c, 0) \rightarrow (c, s)$
$07(S)$		$ACL_Y(f, u, a > 0), (c, s), s \rightsquigarrow K, c = h(K, f)$	$K \rightsquigarrow s', \{f, c, s'\}$
$08(A)$		$AU_Y(f, u, \alpha), (f, 0)$	$(f, 0) \rightarrow (f, \alpha), AU_Y(f, u, \alpha)$
$09(A)$	α'	$AU_X(f, u, \alpha), (f, (u, a)), a > 2$	$(f, \alpha') \rightarrow (f, \alpha), AU_{X,Y}(f, u, \alpha)$
$10(A)$	Z	$VU_X(f, u, \lambda), (f, (u, a)), a > 1$	$VU_Z(f, u, \lambda)$
$11(A)$	$\{f, u\}, Z, a$	$(f, (u, a))$	$ACL_Z(f, u, a)$
$12(A)$	$\{f, u\}$	$(f, 0)$	$ACL_Z(f, u, 0)$
$13(A)$	θ	$ACL_Z(f, 0, 0), (f, \theta)$	$(f, \theta) \rightarrow (f, 0)$
$14(F)$	Z	$VU_Y(f, u, \lambda), (f, 0)$	$(f, 0) \rightarrow (f, (1, \lambda)), FP_Z(f, 1, u, \lambda, 1)$
$15(F)$	Z, q, λ'	$VU_X(f, u, \lambda), (f, (q-1, \lambda')), (f, (q, 0))$	$(f, (q, 0)) \rightarrow (f, (q, \lambda)), FP_Z(f, q, u, \lambda, q)$
$16(F)$	Z, q, λ	$ACL_Y(f, u, a > 0), (f, (q, \lambda)), (f, (q+1, 0))$	$FP_Z(f, q, u, \lambda, q)$
$17(F)$	Z, q, λ	$ACL_X(f, u, a > 0), (f, (q, \lambda))$	$FP_Z(f, q, u, \lambda, 0)$

event 06. To send a secret to a user, event 11 can be invoked to create a ACL message confirming that the user has the requisite access right, to trigger event 07. If the user does not have any access to f, event 11 or 12 can be triggered to convey this fact to the user.

A user u merely requesting file parameters for the latest version of a file can be satisfied by invoking event 11 to create a ACL message that triggers event 16. The preconditions for event 16 ensure that q is the latest version through non-existence of version $q+1$ (pre-condition $(f, (q+1, 0))$). If the user requests a specific (older) version, event 17 is triggered instead. Note that the FP message created this time sets the field corresponding to the highest version number to 0 to indicate that it did not "bother to check" the highest version number. The FP message created by event 16 or 17 triggers event 05. Once again, if the user does not have any access privilege, or the file does not exist, event 11 or 12 can be triggered to convey this fact to the user.

At first sight, it may appear that replay attacks (for example, an old request for ACL update may be replayed), are ignored. This omission is deliberate, as strategies to prevent replays can be addressed by generic protocols for creation and verification of MN messages.

IV. Discussions and Conclusions

A novel mirror network model for securing ISes was outlined, driven by the need to reduce complexity of both the assurance software for any IS, and the platform on which the assurance software is executed. The complexity of the platform was kept low by deliberately constraining MN modules to perform only logical and hash operations. The complexity of the assurance software was minimized by constraining it to be a list of simple pre-conditions and post-conditions.

The only assumptions behind the MN approach are i) the correctness of the MN specification for the IS to be secured, and ii) the integrity of MN modules. No hardware/software of the IS itself need to be trusted to realize the desired assurances. As the MN specification is open, anyone with IS domain knowledge can verify its correctness. As the MN modules are deliberately constrained to possess simple and identical functionality, an infrastructure for mass production (possibly as chips), verification, and certification of MN modules can be realized at a reasonably low cost.

A. Comparison With Trinc

Another approach in the literature which leverages a simple trustworthy module specification to bootstrap system assurances is Trinc [2]. Specifically, a *trinket* is a module following the *Trinc* specification, whose sole purpose is the attestation of monotonic counters stored inside the trinket.

Similar to MN modules, every trinket has a unique identity. Every trinket also has an asymmetric key pair certified against its identity. A primary counter in the trinket is leveraged to create a plurality of secondary counters as follows. Whenever a new secondary counter is created, it is identified by the current value of the primary counter — which is incremented on creation of the new counter. Built-in functions of a trinket can be used to a) request a trinket to create a new counter with identity n, or b) bind (by computing a digital signature) some arbitrary value x and an incremented counter value $c'_n \geq c_n$ (where c_n is the current value of counter with identity n).

As an example, consider a scenario where a dynamic constrained data item (for example, a file hash) F, is bound to a counter with identity n, and value c_n, in a trinket with identity G. More specifically, let a value x bound to the counter (n, c_n, G) (through a signature of trinket G) represent a one way function of the signature of the provider/owner of file F. Whenever F is updated, the owner ensures that a fresh signature x' is bound to $(F, n, c'_n > c_n, G)$ — by requesting the trinket G to update the counter n to c'_n, and issue a certificate binding x' to the updated counter. Anyone receiving the file F (even from an untrusted repository) can verify its freshness by obtaining the attestation by G (binding x' to its current counter (n, c'_n, G)). Specifically, as the counter n is no longer c_n, the old value of F (along with signature x) can *not* be replayed by the repository.

To reduce the overhead associated with digital signatures, the Trinc specification also includes an alternate mechanism to attest/verify certificates, using shared symmetric secrets. In this case, however, a system-specific trusted third party is required to set-up secrets bound to specific counters of different trinkets — which are then shared between all entities that are required to verify the attestation.

Compared to MN modules, the main disadvantages of Trinc are as follows. Firstly, the Trinc specification limits the number of counters that can be "remembered" (and hence, the number of CDIs that can be reliably tracked by a trinket) to a small queue length (10 to 15). One way to overcome this limitation is by addition of built-in Merkle-tree functionality in Trinc for tracking any number of counters using a single monotonic counter [13].

Even with this addition, strategies to secure any practical system using Trinc will, unfortunately, require components *other* than the Trinc modules to be trusted. The reason for this is that Trinc by itself does not offer an explicit mechanism for binding a Trinc identity G to a specific subsystem/database that includes data F, or binding a specific piece of data F associated with the subsystem to a specific counter n in a specific trinket G. In the specific example above, the owner of F is trusted to do so. Unlike the MN model, where such system-specific bindings (to enforce system-specific rules) can be taken into account by the rich MN specification, the Trinc model does *not* have the ability to enforce IS specific rules. The shortcomings of Trinc are addressed by addition of simple (in terms of resource requirements), yet rich functionality to MNs modules — IOMT and mutual authentication capabilities — which require demand simple sequences of hash operations.

B. Conclusions

Current approaches to secure ISes predominantly rely on

1) ever changing *reactionary* measures (software updates, IDSes, firewalls) to improve the integrity of different subsystems and
2) cryptographic strategies for securing interactions between subsystems.

The former strategies are plagued by the possibility of new bugs in updates, and/or bugs in the very design of complex IDSes. Breaches in the latter (cryptographic) strategies [12] often result from the lack of integrity in the environment in which cryptographic protocols are executed (after all, a cryptographic algorithm is at most only as reliable as the platform in which cryptographic keys are stored and the cryptographic algorithm is executed). The novel and holistic MN approach to assure information systems is motivated by the often repeated (and unfortunately just as often ignored) maxim that "complexity is the enemy of security" [1]. The main novelty of the proposed approach stems from applying system integrity models to the assurance protocol of an IS (instead of the IS itself, as in the Clark-Wilson model).

Not withstanding complexity of ISes, rules that govern *how* data should be manipulated by ISes tend to be simple. Security breaches in systems rarely result from incorrect rules. Rather, they result from issues in the process of *implementing* the rules into a working system. This process includes numerous tasks performed during design, deployment and maintenance of the system, possibly by numerous personnel. It is far from practical to be able to assure the integrity of every component and personnel of such a complex process. The crux of the MN model is that it permits us to short-circuit this process to observe "if an IS is indeed abiding by design rules."

It is important to note that the MN approach does *not* obviate the need for measures necessary to root out malicious functionality in IS components, for if such functionality results in illegal modifications to the IS databases, the IS can no longer demonstrate its integrity to its users. In other words, all that the MN approach guarantees is that ISes will not be able to *hide* security violations from users (and other stake-holders) of the IS.

Our ongoing work involves developing MN rules for a wide range of information systems, with the longer term goal of developing a succinct language for expressing preconditions and post-conditions as instructions that can be easily interpreted by resource limited MN modules.

REFERENCES

[1] B. Schneier, "A plea for simplicity: you can't secure what you dont' understand," Information Security, November 1999.

[2] D. Levin, J. R. Douceur, J. R. Lorch, and T. Moscibroda, "Trinc: Small Trusted Hardware for Large Distributed Systems," 6th USENIX Symposium on Networked Systems Design and Implementation (NSDI 09), Boston, MA, April 2009.

[3] D .D. Clark, and D. R. Wilson, "A Comparison of Commercial and Military Computer Security Policies," in Proceedings of the 1987 IEEE Symposium on Research in Security and Privacy (SP'87), May 1987, Oakland, CA; IEEE Press, pp. 184–193.

[4] M. Ramkumar, Symmetric Cryptographic Protocols, Springer, 2014.

[5] V. Thotakura, and M. Ramkumar, "Minimal TCB For MANET Nodes," 6th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2010), Niagara Falls, ON, Canada, September 2010.

[6] S. D. Mohanty, and M. Ramkumar, "Securing File Storage in an Untrusted Server Using a Minimal Trusted Computing Base," First International Conference on Cloud Computing and Services Science, Noordwijkerhout, The Netherlands, May 2011.

[7] A. Velagapalli, S. Mohanty, and M. Ramkumar,"An Efficient TCB for a Generic Data Dissemination System," International Conference on Communications in China: Communications Theory and Security (ICCC-CTS), 2012.

[8] R. C. Merkle, "A Digital Signature Based on a Conventional Encryption Function," Advances in Cryptology, CRYPTO '87. Lecture Notes in Computer Science 293. 1987.

[9] M. Ramkumar, "Trustworthy Computing Under Resource Constraints With the DOWN Policy," IEEE Transactions on Secure and Dependable Computing, pp 49-61, Vol 5, No 1, Jan-Mar 2008.

[10] M. Ramkumar, "The Subset Keys and Identity Tickets (SKIT) Key Distribution Scheme," IEEE Transactions on Information Forensics and Security (TIFS), pp 39-51, Vol 5, No 1, Mar 2010.

[11] M. Ramkumar, "On the Scalability of a "Nonscalable" Key Distribution Scheme," IEEE SPAWN 2008, Newport Beach, CA, June 2008.

[12] Z. Durumeric et. al., "The Matter of Heartbleed," IMC 2014, Vancouver, Canada, Nov 2014.

[13] Sarmenta, L. F. G. Dijk, M. V. ODonnell, C. W. Rhodes, and S. Devadas, "Virtual Monotonic Counters and Count-Limited Objects using a TPM Without a Trusted OS," Proceedings of the 1st ACM CCS Workshop on Scalable Trusted Computing (STC06), pages 27–42, 2006.

An Investigation of a Factor That Affects the Usage of Unsounded Code Strings at the End of Japanese and English Tweets

Yasuhiko Watanabe, Kunihiro Nakajima, Haruka Morimoto, Ryo Nishimura, and Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan
Email: watanabe@rins.ryukoku.ac.jp, t13m071@mail.ryukoku.ac.jp, t13m076@mail.ryukoku.ac.jp,
r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

Abstract—**In this study, we compare Japanese and English tweets submitted to Twitter and discuss how we use unsounded code strings at the end of online messages. We first define unsounded codes and unsounded code strings used in Japanese and English text. Next, we compare and discuss the usage of unsounded code strings at the end of Japanese and English tweets, especially, to general public and particular persons. Finally, we show the receiver of a tweet, whether general public or a particular person, is a factor that affects the usage of unsounded code strings at the end of Japanese and English tweets. Specifically, Japanese speakers use unsounded code strings at the end of tweets more frequently to particular persons than to general public while English speakers do not.**

***Keywords*–unsounded code string; Twitter; general public; particular persons; non verbal communication.**

I. Introduction

Many of us think that it is easy to understand the meanings of non verbal expressions in online messages even if different language speakers generate them. However, the usage differs between different language speakers and the difference is at risk of bringing unnecessary frictions between them. As a result, it is important to consider the difference, especially, in multilingual computer-mediated communication (CMC) systems. In this study, we show the usage of unsounded code strings, one kind of non verbal expression, differs between Japanese and English speakers and discuss a factor that affects the usage of them.

We often find consecutive unsounded marks and characters are used at the end of online messages, such as mails, chattings, and tweets in Twitter.

(exp 1) I'm freezing!!!!
(exp 2) *@ryuuuuuuuu_2012 soushita hou ga iiyo.......*
(@ryuuuuuuuu_2012 you had better do it.......)

(exp 1) and (exp 2) are tweets submitted to Twitter. (exp 1) was submitted by an user who chose English as his/her language for tweets. On the other hand, (exp 2) was submitted by an user who chose Japanese as his/her language for tweets. Both (exp 1) and (exp 2) have consecutive unsounded marks at the end of them. These unsounded marks are used for smooth communication. The submitter of (exp 1) is thought to use the three consecutive exclamation marks for expressing his/her impression strongly. On the other hand, the submitter of (exp 2) is thought to use the seven consecutive periods for expressing his/her opinion softly. In this study, we define unsounded marks and characters as *unsounded codes*. Furthermore, we define three or more consecutive unsounded codes as a *unsounded code string*. For example, in Twitter, 14 % of Japanese tweets and 10 % of English tweets have unsounded code strings at the end of them. Although unsounded code strings are popular, there are few studies on them. As a result, in this study, we investigate how we use unsounded code strings at the end of tweets in Twitter. Especially, we compare Japanese and English tweets in Twitter and discuss a factor that affects the usage of unsounded code strings at the end of tweets. The results of this study will give us a chance to understand the usage of unsounded code strings and improve multilingual CMC systems.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we define unsounded code strings and describes how they are used at the end of tweets in Twitter. Finally, in Section IV, we present our conclusions.

II. Related works

There are a considerable number of studies comparing speakers of various languages from various viewpoints, such as, pragmatics, cognitive science, and so on. These studies can be classified into two types:

- studies comparing native speakers of one language to non-native speakers of the same language
- studies comparing native speakers of one language to native speakers of other language directly

This study is classified into the latter. It is because we compare unsounded code strings in Japanese tweets to those in English tweets directly.

In pragmatics, a considerable number of studies have been made on interlanguage speech acts, such as, expressing compliments [1], apologies [2], gratitude [3], politeness [4], and refusals [5]. In these studies, native speakers of one language were compared to non-native speakers of the same language. Also, there are a considerable number of studies comparing pauses and backchannels of native speakers to those of non-native speakers. Backchannels are listener's responses, such as "uh-huh" and "yeah", given while someone else is talking, to show an interest, attention, or willingness to keep listening. Deschamps investigated how French learners of English made pauses in their English speeches [6]. Bilá and Džambová investigated the function of silent pauses in native and non-native speakers of English and German [7]. Ishizaki investigated how English, French, Chinese and Korean learners of Japanese and native Japanese speakers made pauses in their Japanese speeches [8]. Tavakoli reported that the location of pauses is important in comparisons of native speakers and

TABLE I. The numbers of Japanese and English normal tweets, replies, and retweets (from November/2012 to December/2012).

language	tweet type	normal tweet (%)	reply (%)	retweet (%)	total (%)
Japanese	all	3,813,164 (53.82%)	2,528,642 (35.69%)	743,461 (10.49%)	7,085,267 (100.00%)
	with UCS	356,727 (36.92%)	430,294 (44.54%)	179,166 (18.54%)	966,187 (100.00%)
English	all	16,023,795 (51.27%)	8,267,646 (26.45%)	6,961,800 (22.28%)	31,253,241 (100.00%)
	with UCS	1,121,952 (34.30%)	632,298 (19.33%)	1,516,571 (46.37%)	3,270,821 (100.00%)

”with UCS” means tweets that have unsounded code strings at the end of them

foreign learners [9]. Okazawa reported that native Japanese speakers paused roughly twice as often in their English utterances than in their Japanese utterances [10]. LoCastro reported that Japanese speakers often feel uncomfortable when speaking English because they are unable to use the appropriate backchannels [11]. From the viewpoint of cognitive science, Tera et al. compared and analyzed reading processes of native Japanese speakers and foreign learners of Japanese [12].

On the other hand, there are also a considerable number of studies comparing native speakers of one language to native speakers of other language directly. Especially, many studies have been made on backchannels. It is because backchannels are found in various languages. Maynard showed that backchannel phenomena for Japanese and English differ in terms of type, frequency, and context [13]. Miller reported that Japanese speakers use backchannels more frequently than English speakers [14]. However, in most of previous studies, the frequency of backchannels were observed in various situations while little is known about factors that affect the frequency of backchannels. Chen pointed out that it is important to investigate factors that affect the frequency of backchannels and took White's work [15] for example [16]. White reported that Americans used backchannels more frequently in conversations with Japanese than in conversations with other Americans [15]. In other words, the conversation partner, whether American or Japanese, is a factor that affects the frequency of Americans' backchannels. As a result, when we compare different language speakers, it is important to investigate factors providing different responses between them. In this study, we investigate a factor that affects the usage of unsounded code strings at the end of tweets.

III. Unsounded code strings at the end of Japanese and English tweets in Twitter

In this section, we compare and discuss the usage of unsounded code strings at the end of Japanese and English tweets. In Section III-A, we define unsounded code strings and show how they are used at the end of tweets. In Section III-B, we show the investigation object of this study. Finally, in Section III-C, we compare Japanese and English tweets and discuss a factor that affects the usage of unsounded code strings at the end of tweets.

A. The definition of an unsounded code string

First, we define unsounded codes and unsounded code strings. In this study, we define that an unsounded code string is three or more consecutive unsounded codes. In this study, unsounded codes in English text are limited to

- punctuation marks (e.g. !#$%&.,:;<=>?@ (){}).

On the other hand, unsounded codes in Japanese text are limited to the following marks and characters:

- punctuation marks,
- Greek characters,
- Cyrillic characters, and
- ruled lines.

It is because these marks and characters are generally unsounded when they are used at the end of Japanese sentences.

Next, we show how unsounded code strings are used at the end of tweets. Twitter users often use unsounded code strings in order to enable anyone to understand their tweets clearly and avoid unnecessary frictions with others.

(exp 3) *kadai owattaaaaaaaaaaaaa!!!!!!!!!!* (I got my homework dooooooooooooone!!!!!!!!!!)
(exp 4) @jayne_hurley looks amazing !!!
(exp 5) WEIRDEST dream last night omg...

For example, the submitters used exclamation marks consecutively at the end of (exp 3) and (exp 4) for expressing their feelings strongly. On the other hand, the submitter used periods consecutively at the end of (exp 5) for expressing his/her impression softly.

We may note that there are some submission constraints in Twitter. For example, Twitter users are prohibited to post the same tweets repeatedly. To avoid this constraint, some users use unsounded code strings. For example, (exp 6), (exp 7), and (exp 8) were posted consecutively in a few seconds to a particular user, *ssuzuki16*, beyond the limit of repeated submission.

(exp 6) @*ssuzuki16 yo...*
(exp 7) @*ssuzuki16 yo....*
(exp 8) @*ssuzuki16 yo.....*

B. The investigation object

We obtained tweets by using the streaming API. However, the streaming API allows us to obtain only 1% of all public streamed tweets because of API restriction. We used the streaming API and obtained the following tweets in three weeks in November and December 2012.

- 7,085,267 tweets submitted by users who chose Japanese as their language for tweets. In this study, we call these tweets *Japanese tweets*.
- 31,253,241 tweets submitted by users who chose English as their language for tweets. In this study, we call these tweets *English tweets*.

Table I shows the number of the obtained Japanese and English tweets. These tweets can be classified into three types:

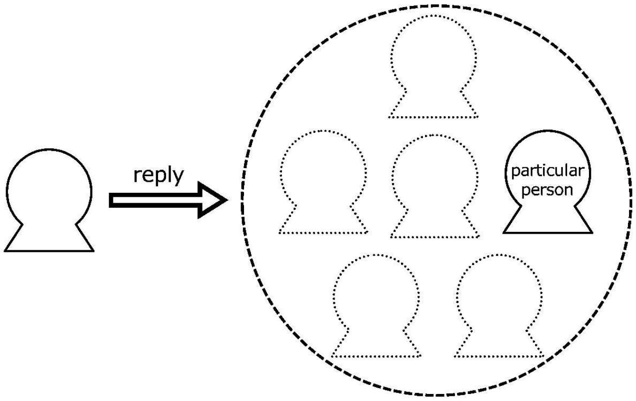


Figure 1. A reply is submitted to a particular user.

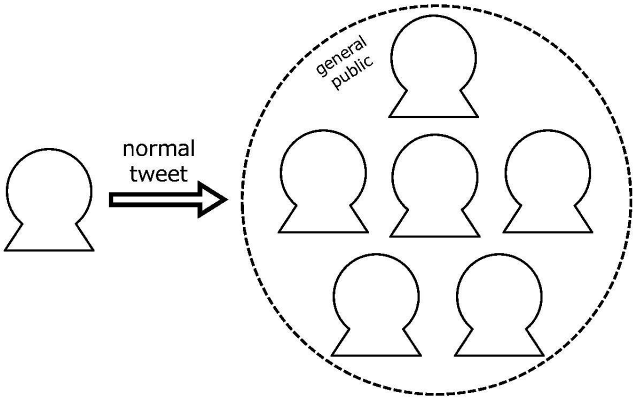


Figure 2. A normal tweet is submitted to general public.

- reply
  A reply is submitted to a particular user (Figure 1). It contains “@username” in the body of the tweet. For example, (exp 2) and (exp 4) are replies.
- retweet
  A retweet is a reply to a tweet that includes the original tweet.
- normal tweet
  A normal tweet is neither reply nor retweet. For example, (exp 1), (exp 3), and (exp 5) are normal tweets. Twitter users generally submit their tweets to general public. As a result, most of normal tweets are submitted to general public (Figure 2).

From the obtained Japanese tweets, we extracted 966,187 Japanese tweets that have unsounded code strings at the end of them. These 966,187 Japanese tweets are 13.64% of all the Japanese tweets. On the other hand, from the obtained English tweets, we extracted 3,270,821 English tweets that have unsounded code strings at the end of them. These 3,270,821 English tweets are 10.47% of all the English tweets. Table I shows the number of Japanese and English normal tweets, replies, and retweets that have unsounded code strings at the end of them.

In this study, we do not discuss unsounded code strings at the end of retweets. It is because, messages in retweets are created not by submitters, but by other users. As a result, retweets are inadequate to investigate how we use unsounded code strings at the end of online messages.

In this study, we compare unsounded code strings at the end of (1) normal tweets and (2) replies. It is because we intend to compare and discuss the usage of unsounded code strings at the end of tweets to (1) general public and (2) particular persons. As mentioned, normal tweets are generally submitted to general public. On the other hand, each reply is submitted to a particular person.

Figure 3 and Figure 4 show the cumulative relative frequency distribution of

- the length of all the Japanese and English tweets (excluding retweets),
- the length of Japanese and English tweets (excluding retweets) that have unsounded code strings at the end of them, and
- the length of unsounded code strings at the end of Japanese and English tweets (excluding retweets).

As shown in Figure 3 and Figure 4, the distribution of the length of Japanese tweets shifts to shorter ranges than the length of English tweets. On the other hand, the distribution of the length of unsounded code strings at the end of Japanese tweets shifts to longer ranges than the length of those at the end of English tweets.

C. The comparison of the usage of unsounded code strings at the end of Japanese and English tweets

Next, we compare the length of unsounded code strings at the end of normal tweets and replies. Figure 5 and Figure 6 show the cumulative relative frequency distribution of the length of unsounded code strings at the end of

- Japanese normal tweets and replies, and
- English normal tweets and replies.

As shown in Figure 5 and Figure 6, the length of unsounded code strings at the end of Japanese and English normal tweets have a similar distribution pattern to those of Japanese and English replies, respectively. As a result, it may be said that the length of unsounded code strings at the end of tweets are less affected by whether the tweets are normal tweets or replies.

Next, we compare the length of normal tweets and replies that have unsounded code strings at the end of them. Figure 7 and Figure 8 show the cumulative relative frequency distribution of

- the length of all the Japanese and English normal tweets, and
- the length of Japanese and English normal tweets that have unsounded code strings at the end of them.

On the other hand, Figure 9 and Figure 10 show the cumulative relative frequency distribution of

- the length of all the Japanese and English replies, and
- the length of Japanese and English replies that have unsounded code strings at the end of them.

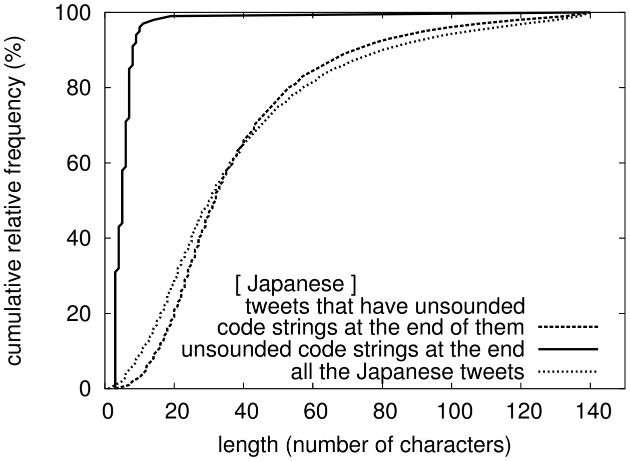


Figure 3. The cumulative relative frequency distribution of the length of (1) all the Japanese tweets, (2) Japanese tweets that have unsounded code strings at the end of them, and (3) unsounded code strings at the end of them.

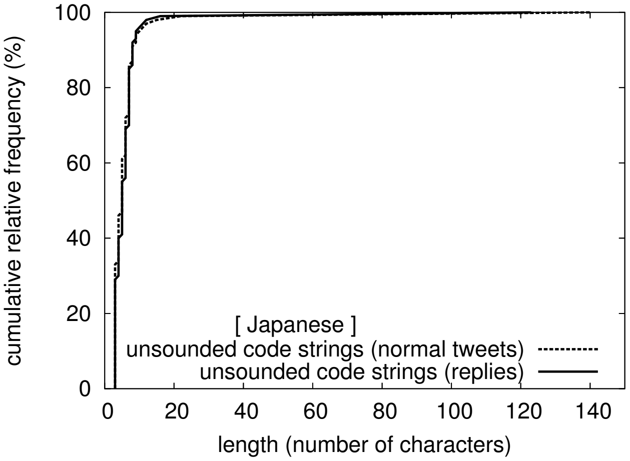


Figure 5. The cumulative relative frequency distribution of the length of unsounded code strings at the end of Japanese normal tweets and replies.

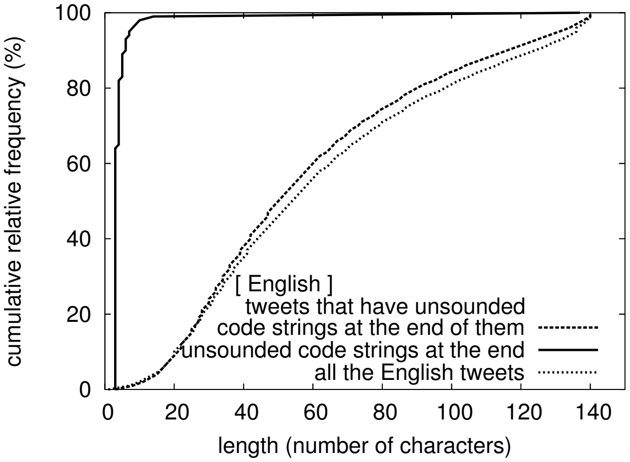


Figure 4. The cumulative relative frequency distribution of the length of (1) all the English tweets, (2) English tweets that have unsounded code strings at the end of them, and (3) unsounded code strings at the end of them.

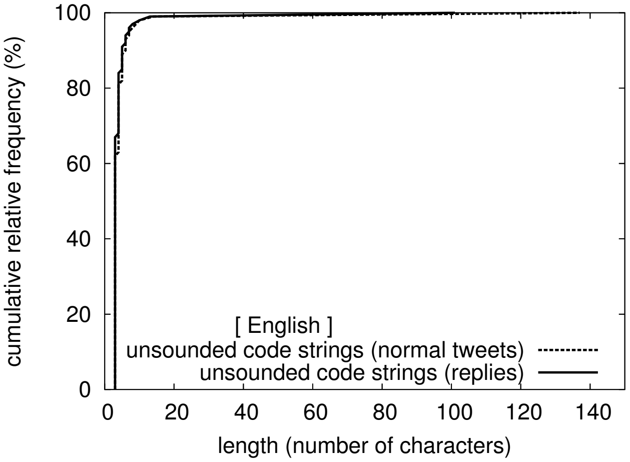


Figure 6. The cumulative relative frequency distribution of the length of unsounded code strings at the end of English normal tweets and replies.

As shown in Figure 7–10, only the distribution of the length of Japanese replies that have unsounded code strings at the end of them shifts to longer ranges than the length of all the Japanese replies. On the other hand, the distribution of the length of the other tweets (Japanese normal tweets, English normal tweets and replies) that have unsounded code strings at the end of them do not shift to longer ranges when they are longer than 30 characters. It may be said that the length of Japanese tweets that have unsounded code strings at the end of them are affected by whether the tweets are normal tweets or replies. On the other hand, the length of English tweets that have unsounded code strings at the end of them are not affected.

Next, we investigate that the percentages of tweets that have unsounded code strings at the end of them are affected by whether the tweets are normal tweets or replies. Figure 11 shows the percentages of Japanese normal tweets and replies that have unsounded code strings at the end of them. As shown in Figure 11, 9.4 % of Japanese normal tweets have unsounded code strings at the end of them while 17.0 % of Japanese replies have unsounded code strings at the end of them. As a result, the percentages of Japanese normal tweets and replies that have unsounded code strings at the end of them differ considerably from each other. In other words, Japanese replies have unsounded code strings at the end of them more frequently than Japanese normal tweets. On the other hand, Figure 12 shows the percentages of English normal tweets and replies that have unsounded code strings at the end of them. As shown in Figure 12, 7.0 % of English normal tweets have unsounded code strings at the end of them while 7.6 % of English replies have unsounded code strings at the end of them. As a result, the percentages of English normal tweets and replies that have unsounded code strings at the end of them differ little from each other. In addition, the percentages of English normal tweets and Japanese normal tweets that have

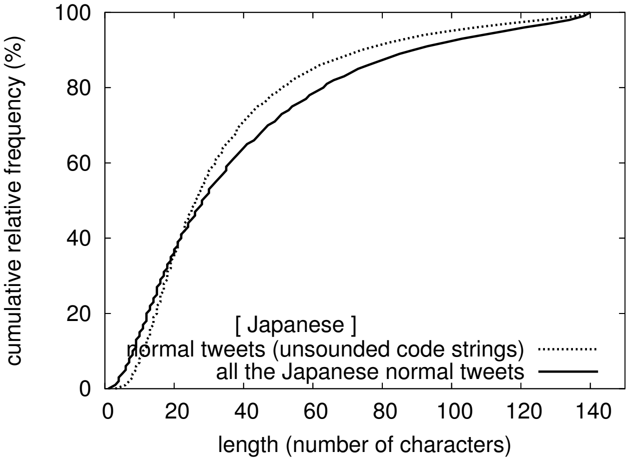


Figure 7. The cumulative relative frequency distribution of the length of all the Japanese normal tweets and Japanese normal tweets that have unsounded code strings at the end of them.

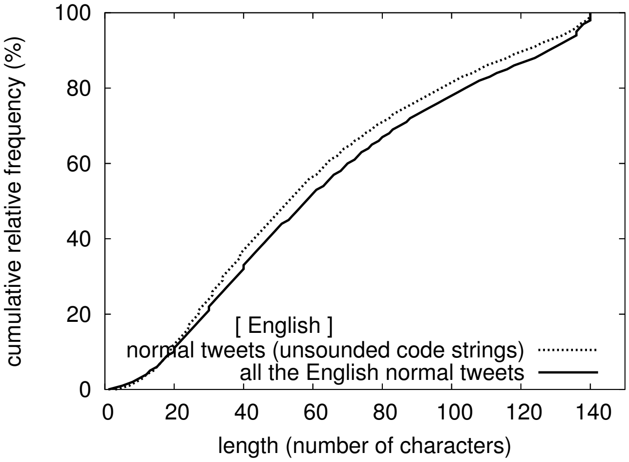


Figure 8. The cumulative relative frequency distribution of the length of all the English normal tweets and English normal tweets that have unsounded code strings at the end of them.

unsounded code strings at the end of them differ little from each other. From these points, it may be said that Japanese speakers use unsounded code strings at the end of tweets more frequently to particular persons than to general public while English speakers do not. In other words, the receiver of a tweet, whether general public or a particular person, is a factor that affects the usage of unsounded code strings for Japanese speakers, however, not for English speakers. It is not clear whether this phenomenon is specific for Japanese speakers.

IV. Conclusion

Unsounded code strings, in other words, consecutive unsounded marks and characters are frequently used at the end of online messages. However, there were few studies on them. In this study, we investigated unsounded code strings at the end of Japanese and English tweets in Twitter. Then, we showed that Japanese speakers use unsounded code strings at the end of tweets more frequently to particular persons than to general public while English speakers do not. It may be said that the receiver of a tweet, whether general public or a particular person, is a factor that affects the usage of unsounded code strings for Japanese speakers, however, not for English speakers. In order to discuss whether this phenomenon is specific for Japanese speakers, we intend to analyze tweets in various languages. The results of this study will give us a chance to understand the usage of unsounded code strings and improve multilingual CMC systems.

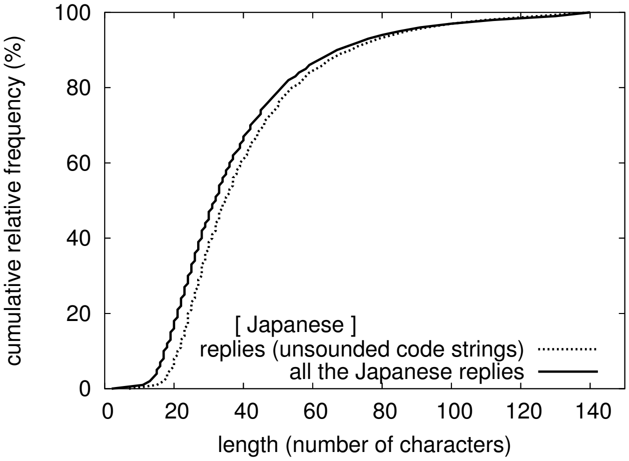


Figure 9. The cumulative relative frequency distribution of the length of all the Japanese replies and Japanese replies that have unsounded code strings at the end of them.

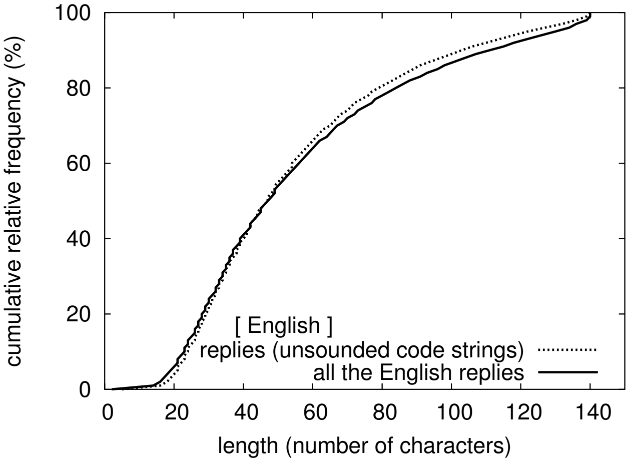


Figure 10. The cumulative relative frequency distribution of the length of all the English replies and English replies that have unsounded code strings at the end of them.

References

[1] N. Wolfson, "The social dynamics of native and nonnative variation in complimenting behavior," in The Dynamic Interlanguage: Empirical Studies in Second Language Variation. Springer US, 1989, pp. 219–236.

[2] M. L. Bergman and G. Kasper, "Perception and performance in native and nonnative apology," in Interlanguage pragmatics. Oxford University Press, 1993, pp. 82–107.

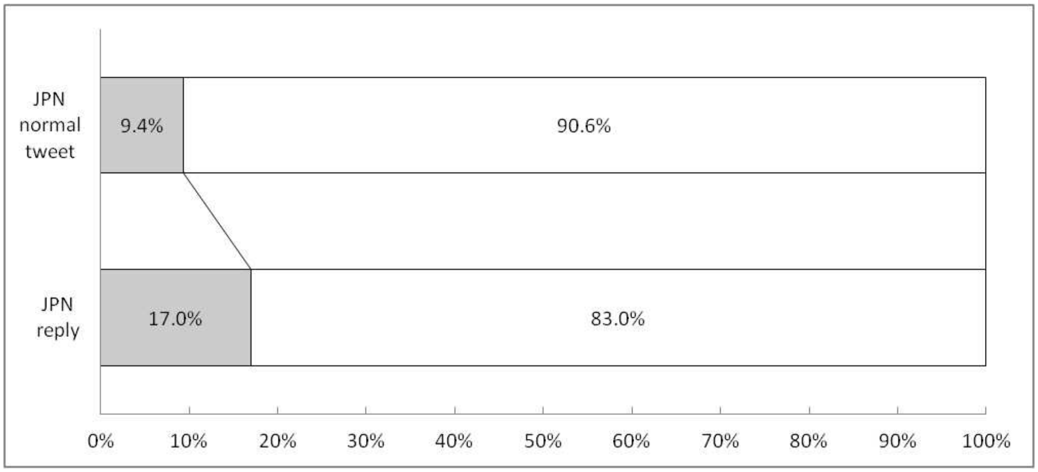


Figure 11. The percentages of Japanese normal tweets and replies that have unsounded code strings at the end of them

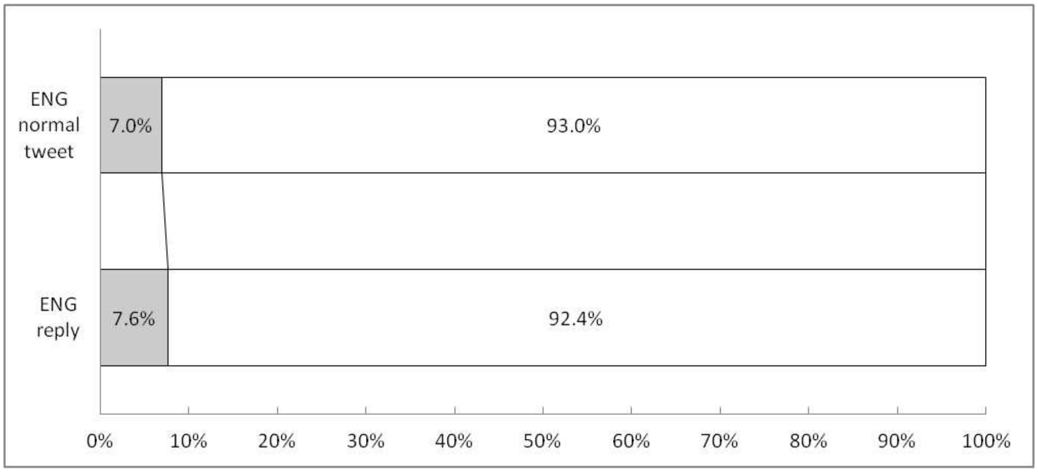


Figure 12. The percentages of English normal tweets and replies that have unsounded code strings at the end of them

[3] M. Eisenstein and J. Bodman, "Expressing gratitude in American English," in Interlanguage pragmatics. Oxford University Press, 1993, pp. 64–81.

[4] S. Tanaka and S. Kawade, "Politeness strategies and second language acquisition," Studies in Second Language Acquisition, vol. 5, no. 1, 1982, pp. 18–33.

[5] L. M. Beebe, T. Takahashi, and R. Uliss-Weltz, "Pragmatic transfer in ESL refusals," in Developing communicative competence in a second language. Newbury House Publishers, 1990, pp. 55–73.

[6] A. Deschamps, "The syntactical distribution of pauses in English spoken as a second language by French students," in Temporal variables in speech. De Gruyter Mouton Publishers, 1980, pp. 255–262.

[7] M. Bilá and A. Džambová, "A preliminary study on the function of silent pauses in L1 and L2 speakers of English and German," Brno Studies in English, vol. 37, no. 1, 2011, pp. 21–39.

[8] A. Ishizaki, "How does a learner leave a pause when reading Japanese aloud?: A comparison of English, French, Chinese and Korean learners of Japanese and native Japanese speakers (in Japanese)," Japanese-Language Education around the Globe, vol. 15, 2005, pp. 75–89.

[9] P. Tavakoli, "Pausing patterns: differences between L2 learners and native speakers," ELT Journal, vol. 65, no. 1, 2011, pp. 71–79.

[10] S. Okazawa, "Pauses and fillers in second language learners' speech," Studies in Language and Culture, vol. 23, 2014, pp. 52–66. [Online]. Available: http://opac.library.twcu.ac.jp/opac/repository/1/5697/ [retrieved: August, 2015]

[11] V. LoCastro, "Aizuchi: A Japanese conversational routine," in Discourse Across Cultures: Strategies in World Englishes. Prentice Hall, 1987, pp. 101–113.

[12] A. Tera, K. Shirai, T. Yuizono, and K. Sugiyama, "Analysis of eye movements and linguistic boundaries in a text for the investigation of Japanese reading processes," IEICE Transactions on Information and Systems, vol. E91.D, no. 11, 2008, pp. 2560–2567.

[13] S. K. Maynard, "On back-channel behavior in Japanese and English casual conversation," Linguistics, vol. 24, no. 6, 1986, pp. 1079–1108.

[14] L. Miller, "Verbal listening behavior in conversations between Japanese and Americans," in The Pragmatics of Intercultural and International Communication. John Benjamins Publishing, 1991, pp. 110–130.

[15] S. White, "Backchannels across cultures: A study of Americans and Japanese," Language in Society, vol. 18, no. 1, 1989, pp. 59–76.

[16] C. Tzuching, "Existing research of Japanese backchannels : An overview for the future (in Japanese)," Japanese language education, vol. 2002, 2002, pp. 222–235.