Passive Packet Loss Detection in Wi-Fi Networks and its Effect on HTTP Traffic Characteristics

Saeed Ullah⋆, Imdad Ullah•, Hassaan Khaliq Qureshi•, Rim Haw⋆, Sungman Jang⋆, and Choong Seon Hong⋆
⋆ Department of Computer Engineering, College of Electronics and Information,
Kyung Hee University, South Korea
• School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST), Islamabad, Pakistan.
Email: {saeed, rhaw, smjang, cshong}@khu.ac.kr, {hassaan.khaliq, imdad.ullah}@seecs.edu.pk

***Abstract*—In this paper, we analyze the performance of web applications that run on top of TCP over the passively collected data from an operational network based on wireless infrastructure. In particular, we evaluate the impact of 1) TCP loss rates; 2) user interrupts; 3) connection volume and their durations; and 4) variations in TCP throughputs on web applications. Since TCP behavior is dependent on events like packet losses, timeouts, and connection terminations, which affect the behavior of TCP-based web applications. The phenomena is more prominent in wireless networks. Our collected real network traffic traces provides very interesting insight on the TCP behavior showing the effect of TCP throughput on packet losses and user interrupted connections, which in turn also increases unnecessary retransmissions. Our performance evaluation reveal that the longer duration and higher volumes of TCP connections is a function of higher interrupted connections and loss rates. By further studying the statistical properties of TCP connections shows that the loss rate increases with decrease in connection duration, and also leads to frequent termination of TCP connections.**

***Index Terms*—Web Traffic Measurement, Packet Loss, Wi-Fi Networks, Network Analysis, TCP.**

I. Introduction

Data analysis has been widely used to predict different characteristics of computer networks and its applications [1] [2] [3]. Passive measurements, over the data collected within the operational networks, is an effective way of understanding the dynamics of the network traffic. In addition, it has impact on network entities and more importantly on web-applications. Characterizing the traffic through the passive measurements, as opposed to the active measurements, is focused for a number of reasons like 1) the active traffic analysis makes it difficult to better facilitate the packet inspection in order to distinguish different kinds of packets and effectively analyze them for different events of congestion, longer delay etc; 2) the protocol complexity being analyzed, such as the TCP, deeply requires the knowledge of the protocol's internal details through the use of deep packet inspection; 3)the traffic dynamics and its behavior can be analyzed only if the forward packets and their respective response packets are analyzed in cooperation, which is even harder to detect the traffics with the interrupted flows and the unidirectional traffic flows, such as UDP flows.

TCP is a dominant transport protocol used by a wide range of applications such as the web-services, email services and newly emerging peer to peer applications. Understanding the behavior of TCP and its performance limiting factors, such as congestion events and intermediate nodes' buffer limitations, is currently a great interest in order to effectively deploy reliable services and traffic shaping. TCP exploits sequenced segments and acknowledgements to detect and react to the traffic losses. For similar purpose, a sliding window mechanism is used for the error recovery and congestion control mechanisms [4]. Moreover, the TCP recognizes the segment loss as a congestion notification event where the sending rate is reduced and the communicating parties experience increased delays. Due to these reasons, quantifying the impact of the events is essential for investigating the performance of TCP.

For the reasons mentioned above, the main objective of this paper is to analyze the performance of the TCP connections in Wi-Fi networks, since the wireless medium is more prone to errors [5], consequently this study gives a good insight into the performance of TCP connections. We are specifically interested in TCP based web applications where the connections are classified as the TCP connections with longer duration and shorter duration. In this way, the TCP connections are more frequently created and terminated while providing a better insight to investigate the performance of TCP. Compared to the previous studies, to the best of our knowledge, it is the first study to evaluate the performance of TCP on wireless networks and provide the reader and researchers an extended set of measurements on huge amount of data passively collected from an operational network in different periods of time. An efficient model is developed for evaluating the performance of TCP with different flow distributions and observe the effect of various loss rates for various TCP connection durations of bi-directional flows. Later, TCP connections are further distinguished as the user interrupts under different loss rates and connection durations. In addition, the throughputs achieved by the TCP connections are determined to effectively evaluate the successful establishment and termination of TCP connections.

Our main contributions are: 1) to develop an efficient module for evaluating the performance of TCP for web traffic; 2) to determine the impact of TCP for shorter and longer duration in Wi-Fi networks; 3) to differentiate user interrupts inferring dissimilar loss ratios; 4) to collect large amount of network traces from operational network.

The rest of the paper is organized as follows. In Section II, we review some relevant literature and describe the motivation behind this work. Section III, presents the proposed methodology, where we discuss the simulation environment including the experimental setup and the loss analysis module. In Section IV, we discuss the performance of web traffic in terms of loss rates, user interrupts, and throughput for various connection durations and flow distributions. We finally conclude the paper in Section V.

II. Background and Motivation

Extensive literature is available on the passive analysis of internet traffic, which address packet loss, pertaining to the network tomography [6] [7] [8] [1]. In addition, it is mainly concerned to inference the network characteristics, in order to identify root causes of connection failures. Moreover, majority of the root causes of transmission failures are packet losses and reordering of packets [4]. In [6], the authors provide techniques which require the end nodes to cooperate by time-stamping the packets to detect shared congested links. Another work is presented in [7] to identify lossy links by comparing various methods of Bayesian inference, random sampling and linear programming to determine the link loss rates and to identify bottleneck links.

The TCP loss and reordering based on different IP identifier fields are investigated in [8] to understand the end-to-end TCP performance. Similarly, the authors in [1] determine the inferring loss rates where a profile-likelihood-based inference approach is used that infers the link loss rates for various topologies. Besides, some other studies in [9] [10] [11] provide TCP-based framework for passive measurement analysis. However, they analyze TCP behavior by observing the TCP data packets and corresponding acknowledgements at the measurement points. These frameworks are based on inferring the packet losses in different events of timeouts, retransmission, and acknowledgement packet loss. Another set of related studies in [12] [13] [2], estimates the packet loss ratio of the TCP connections in the backbone networks, For this purpose, they use passive measurements before and after the measurement points deployed inside the network.

Rapid improvements in the wireless networks technologies (such as 3G, EDGE, GPRS, and WLAN) and mobile terminals (such as smart phones, and PDAs) have significantly changed the way of accessing information worldwide. In addition, the Wi-Fi hotspot deployments are now prevalent across public places such as coffee shops, airports, hospitals, and campus areas, where mobile wireless clients frequently access the web services. Secondly, web applications enable the ease of access and increased availability of information by means of providing web services. At the same time, they also increase the productivity and exploitation in the corporate and educational institutions. Most of the web applications, for instance accessing the World Wide Web, frequently requires the TCP connections to be established and terminated. Therefore, the TCP performance undergoes highly degraded performance for shorter duration of time. The performance of TCP based web applications is further degraded in Wi-Fi networks, because of the lossy nature of wireless medium. For the performance analysis of web applications, we investigated operational networks based on wired infrastructure using the passive analysis of a large amount of data in [3] [14]. However, in this paper, the performance of TCP based web applications are analyzed for Wi-Fi networks in terms of loss rates, throughput, user interrupts, connections volume and duration.

III. Network Traffic Dataset

In this section, we present the data collection infrastructure in an operational network and then we show some preliminary traffic statistics of the collected traces.

A. Data Collection Setup

The passive traces for web traffic is collected from our campus edge router. The whole university campus is divided in sub-systems where each sub-system consist of multiple wireless and wired LANs. Each wireless LAN is connected to the wired closed via 3Com-4500G switches. Similarly, all LAN switches are directly connected to central distribution switch via fibre lines. For the data collection, a physical hardware port were mirrored on the distribution switch (edge router) and data were collected for 3 months without any break. All the incoming and outgoing traffic passes into the edge router. Figure 1 shows the network structure comprises of wireless internet connectivity. The measurement point is connected to the ingress-egress router connected to the network where all the traffics including the web traffic access the internet. Moreover, the wireless network facility covers various departments and different types of users. For instance, only our computer science department building have separate wireless LANs for research labs, faculty offices, administrative and clerical offices, cafeteria, library, and various R&D offices and labs of our industry partners. In consequence, the users (including students) are busy accessing the web traffic most the time that provides a better way of investigating the TCP performance for web traffic.

For the sake of saving space and efficient measurements, we captured first 96 bytes of each packet that contains the packet's headers including the IP and TCP headers and a small portion of data. The sample traces collected over three months (i.e. Sept-Nov 2011) consist of 617 million data packets amounts to $1.75TB$ of data in size intended for web traffic only.

B. Preliminary Traffic Statistics

Table I shows some preliminary statistics of the collected dataset. Average per day statistics show that number of flows are around 50 thousand, number of packets are 68 millions (1352 packets per flow) and per day volume is 20GB (400KB per flow).

C. Packet Loss Estimation

To evaluate the performance of TCP intended for web traffic, we use the Tcpdump [15] to passively collect the traces. These traces are subject to the TCP bi-directional flows that are successfully established between the end points.

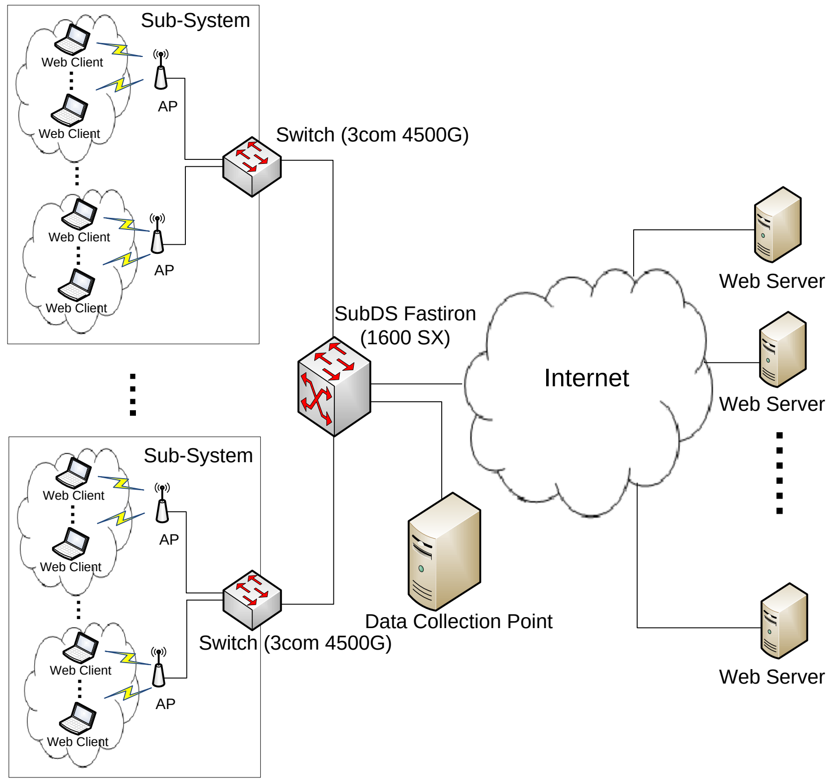

Fig. 1. Architecture of our university campus wireless network

TABLE I
SUMMARY OF TRACES IN VARIOUS MONTHS

	September	October	November	Per Day
Number of Flows	1521731	1557766	1485696	50724
Number of Packets	2058809521	2261969487	1855649555	68626984
Volume (KB)	613313821	636519980	584931318	20386279

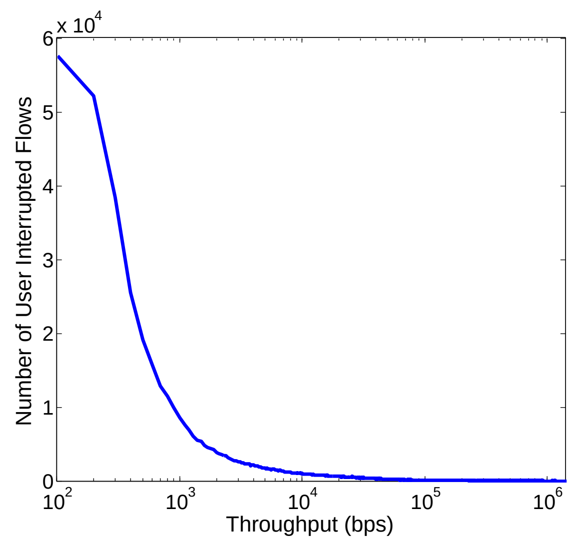

Fig. 2. Client interrupted connections with respect to throughput

Further, to investigate the performance of TCP connections in terms of loss rates, estimation of connection duration, throughput etc, we use the Tstat network traffic analysis tool [16] [17] that can infer detailed information about successful and unsuccessful TCP connections. We use the flow level of TCP session's identification to identify the TCP packet while using 5 different parameters inside the packets, such as the source and destination IP addresses, their port numbers and the protocol number. We then structure the TCP connection using the TCP sequence numbers and their corresponding acknowledgements.

A TCP connection is considered complete when the connections are established using the TCP handshaking (SYNs-ACKs) procedure followed by exchange of data packets, not necessarily, and then followed by the FIN packet. On other hand, the connections terminated with the RST sent from the client side are calculated as the TCP connections interrupted by the clients. The duration of the TCP connection with a particular sequence number is calculated as the start time of the SYN packet and time when either a TCP packet with FIN or RST is indicated. Furthermore, to estimate the average RTT of a connection we use the methodology presented in [4]. However, we ignore connections having less than 4 valid RTT samples, mean the RTTs that are not a result of retransmitted packets in a connection, for eliminating chances of having single erroneous values.

TCP declares a packet loss if it do not receive ACK packet

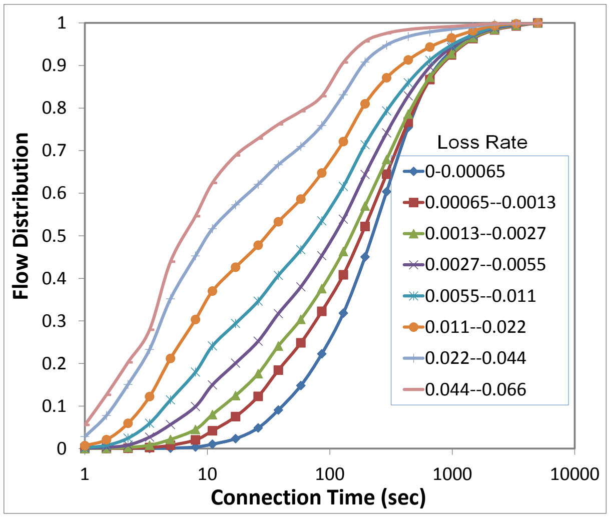

Fig. 3. Effect of Different Loss Rates on Connection Duration with Loss Rates between 10^{-4} and 10^{-2}

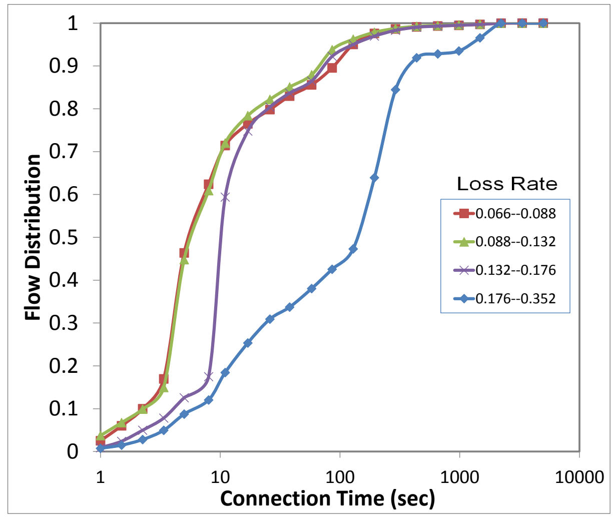

Fig. 4. Effect of Different Loss Rates on Connection Duration with Loss Rates between 10^{-2} and 10^{-1}

for it during Retransmit TimOut (RTO) period or receive 3 duplicate ACKs. To estimate packet loss at the data collection point we have used the method used in [3].

IV. Measurements and Results

We now evaluate the performance of TCP-based web applications over the collected wireless traces. We first analyze the effect of interrupted flows on flow throughput. We then evaluate the effect of various loss rates (categorized as lower and higher loss rates) over the connection duration. The effect of loss rates is also evaluated over the connection volume in terms of successful establishment and termination of TCP connections. Finally, we analyze the effect of loss rates on throughputs, user interrupts.

Fig. 2 show the throughput and interruptions frequencies of those connections that are interrupted from the client side i.e., intentionally terminated connections. the figure shows that low throughput connections are terminated by the users more frequently. This is due to the fact that TCP packets are lost and are retransmitted between the end connections because of interrupting normal TCP flow behavior. Conversely, the average throughput increases as the network observe less number of interrupted flows. As an example, with roughly $5.8 * 10^4$ interrupted flows, the average throughput of TCP is approximately $10^2 bps$ which increases to $10^6 bps$ when interrupts are fewer. An imperative reason for higher number of user interrupts may be due to the impatience of users to wait for the entire duration of successfully downloading the web pages. In the subsequent subsections, we examine the effect of user interrupts over the connection duration, the corresponding throughputs obtained and connection volume for various ranges of loss rates.

A. *Effect of Losses on Connection Duration*

Fig. 3 and Fig. 4 show the effect of connection duration on flow distribution. We note that most of the connections remain active for within $1000 - 5000$ seconds as long as the loss rate remains within the range 10^{-4} and 10^{-2}. Conversely, with the higher loss rates, the connections become shorter in duration and gradually shift to left with less duration of time. We examine that with the loss range between 10^{-2} to 10^{-1}, the duration of the flows remain close between 1 to 100 seconds. This infers a relative behavior between the loss rate and the corresponding connection duration. However, even with the higher loss rates of TCP connection failures, the TCP connection still capable of maintaining the flows sessions to 1000 seconds. There may be several reasons to give clear details of reduction to TCP connection time in case of bad performance, such as the client's self-censorship [2]. Similarly, the shorter duration of TCP connections with higher loss rates may be due to the parallel connection establishment to download various elements of a web page.

B. *Effect of Loss Rate on Connection Size and Duration*

Fig. 5 shows effects of different loss rates on connection's volume and duration. Connection duration is calculated as the time from connection establishment to successful termination of the connection while connection volume is calculated as the amount of total data downloaded by the TCP connections. We observe that with increase in loss rates, the connection duration and the connection volume decreases. This may likely be the fact that the connections are frequently interrupted which also experiences less throughputs. On the contrary, when the loss rate crosses 10% of active TCP connections, the connection time as well as the total number of packets downloaded are abruptly increased. Since TCP is a reliable end-to-end protocol where packet are retransmitted on losses which in turn also increases the connection duration. The connection duration of TCP increases when either the packets are lost or it experiences 3 duplicate ACKs or when Retransmission Time-Out (RTO) expires. The longer delays are further increased by the wireless medium, since the wireless medium observes

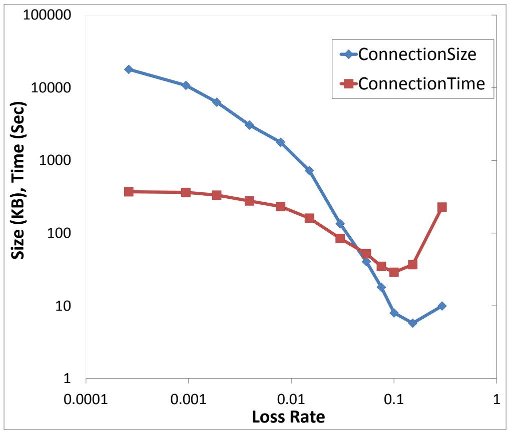

Fig. 5. Effect of loss rate on connection size and duration

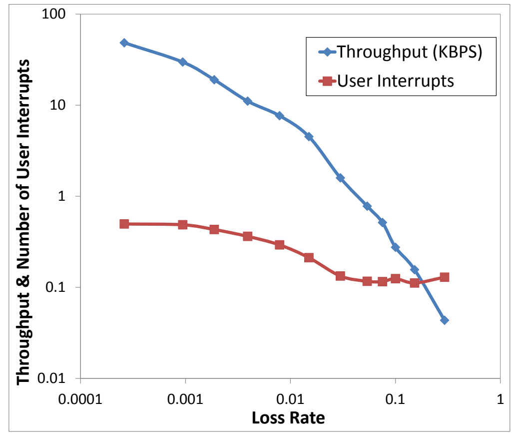

Fig. 6. Throughput and User Interrupts Against Loss Rate

higher losses due to the noisy environment.

C. Effect of Loss Rate on User Interrupts and Flow Throughput

We now analyze the user interrupts and flow throughputs. Fig. 6 shows the throughput and user interruptions for different loss rates. It can be examined that the throughput decreases as the loss rates are increased. For instance, with the loss rate of 10^{-3}, the average throughput obtained is $29Kbps$ which reduces to $0.27Kbps$ with an average loss rate of 10^{-1}. Similarly, the average number of user interrupts fluctuates between $0.1 - 1$.

V. CONCLUSION

To the best of our knowledge, the performance analysis of web applications over Wi-Fi networks had not been investigated. In this paper, we analyzed the effect of various parameters (like TCP loss rates, user interrupts, per connection volume and their duration, TCP throughput) on the performance of TCP-based web applications. Our passive analysis of very large wireless dataset collected over several months shows that the performance of TCP in terms of obtained throughput is significantly affected by loss rates, user interrupted connections, and un-necessary retransmissions. We note that the TCP connection with longer duration and high volumes of data is highly influenced by the loss rates and interrupted connections. We deduce that at the loss rate of higher than 10% of the overall connections established, the TCP connections duration increases. On the other hand, with lower amount of loss rates, the TCP connections are maintained for lesser amount of time and are effectively terminated.

AKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013- H0301-13-4006) supervised by the NIPA(National IT Industry Promotion Agency). Dr. CS Hong is the corresponding author.

REFERENCES

[1] J. Cao, A. Chen, and P. Lee, "Modeling time correlation in passive network loss tomography," in *Dependable Systems & Networks (DSN), 2011 IEEE/IFIP 41st International Conference on*. IEEE, 2011, pp. 550–561.

[2] D. Collange and J. Costeux, "Correlation of packet losses with some traffic characteristics," *Passive and Active Network Measurement*, pp. 233–236, 2007.

[3] I. Din, N. Saqib, and A. Baig, "Passive analysis of web traffic characteristics for estimating quality of experience," in *GLOBECOM Workshops, 2008 IEEE*. IEEE, 2008.

[4] M. Mellia, M. Meo, L. Muscariello, and D. Rossi, "Passive analysis of tcp anomalies," *Computer Networks*, vol. 52, no. 14, pp. 2663–2676, 2008.

[5] D. Bantz and F. Bauchot, "Wireless lan design alternatives," *Network, IEEE*, vol. 8, no. 2, pp. 43–53, 1994.

[6] D. Rubenstein, J. Kurose, and D. Towsley, "Detecting shared congestion of flows via end-to-end measurement," *IEEE/ACM Transactions on Networking (TON)*, vol. 10, no. 3, pp. 381–395, 2002.

[7] V. Padmanabhan, L. Qiu, and H. Wang, "Server-based inference of internet link lossiness," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 1. IEEE, 2003, pp. 145–155.

[8] E. Brosh, G. Lubetzky-Sharon, and Y. Shavitt, "Spatial-temporal analysis of passive tcp measurements," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 2. IEEE, 2005, pp. 949–959.

[9] M. Allman, W. Eddy, and S. Ostermann, "Estimating loss rates with tcp," *ACM SIGMETRICS Performance Evaluation Review*, vol. 31, no. 3, pp. 12–24, 2003.

[10] K. Leung, V. Li, and D. Yang, "An overview of packet reordering in transmission control protocol (tcp): problems, solutions, and challenges," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 18, no. 4, pp. 522–535, 2007.

[11] M. Mellia, M. Meo, L. Muscariello, and D. Rossi, "Passive identification and analysis of tcp anomalies," in *Communications, 2006. ICC'06. IEEE International Conference on*, vol. 2. IEEE, 2006, pp. 723–728.

[12] P. Benko and A. Veres, "A passive method for estimating end-to-end tcp packet loss," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, vol. 3. IEEE, 2002, pp. 2609–2613.

[13] S. Jaiswal, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley, "Measurement and classification of out-of-sequence packets in a tier-1 ip backbone," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*. ACM, 2002, pp. 113–114.

[14] I. Din and N. Saqib, "Passive packet loss detection and its effect on web traffic characteristics," in *Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on*. IEEE, 2008, pp. 729–733.

[15] "Tcpdump's Homepage, http://www.tcpdump.org."

[16] M. Mellia, R. Lo Cigno, and F. Neri, "Measuring ip and tcp behavior on edge nodes with tstat," *Computer Networks*, vol. 47, no. 1, 2005.

[17] "Tstat's Homepage, http://tstat.tlc.polito.it."