Multiadaptive Sampling for Lightweight Network Measurements

João Marco C. Silva and Solange Rito Lima
Departamento de Informática, Centro Algoritmi
Universidade do Minho
Braga, Portugal
Email: id3993@alunos.uminho.pt, solange@di.uminho.pt

***Abstract*—Facing the huge traffic volumes involved in today's networks it is of utmost importance to deploy efficient network measurement solutions to assist network management and traffic engineering tasks correctly, without interfering with normal network operation. Sampling techniques contribute effectively for this purpose as the amount of traffic processed is reduced, ideally without endangering the accuracy of network statistical behavior estimation. Although recent proposals of sampling techniques tend to improve the correctness of the estimation process, their underlying overhead is yet considerably when handling high traffic volumes.**

This paper proposes a new traffic sampling technique for performing lightweight network measurements. This technique, based on linear prediction, is multiadaptive regarding the packet sampling process, allowing to reduce significantly the amount of traffic under analysis while maintaining the representativeness of network samples for accurate network parameters' estimation. The performance evaluation of the sampling technique demonstrates the effectiveness and versatility of the proposal when considering real traces representing distinct traffic load scenarios. The statistical analysis provided evinces that the present solution outperforms classic sampling techniques, both in accuracy and amount of data involved in the measurement process.

I. Introduction

Nowadays, network measurement techniques have to deal with massive traffic volumes which, in many cases, have to be processed online to provide feedback to real-time management and traffic engineering tasks. To operate properly, these tasks have to rely on a accurate view of the status of the network and provided services. Therefore, key aspects to balance when designing an efficient network measurement strategy are the estimation accuracy and the measurement overhead.

Traffic sampling techniques are extensively used to reduce the impact of performing traffic measurements in operational equipment. Their main objective is to select a subset of packets which will be used to estimate network parameters correctly, avoiding processing all network traffic [1].

Existing sampling techniques may be distinguished according to the methodology adopted to select the packets that will integrate a traffic sample. In systematic sampling, a deterministic packet selection function is used, based on (i) the packet position (count-based); (ii) the packet arrival time to the measurement point (time-base); or (iii) the packet contents (content-based). Despite its simplicity, the traffic pattern associated with deterministic sampling may still be heavy to measurement points and produce biased samples.

Random sampling techniques try to avoid biasing the samples by selecting packets according to a random function [2]. Nevertheless, these techniques cannot be deployed to estimate multipoint metrics, such as edge-to-edge delay, since the sampling processes on the two measurement points involved are not correlated and there are no guarantees that samples will be constituted by the same packets [3].

In adaptive sampling techniques the packet selection process considers the value of a reference parameter observed during the measurement period. In this way, the sampling process is more flexible and self-adaptive, and packet selection may change dynamically during the measurement period.

Despite the evolution of adaptive techniques in estimating network performance parameters correctly, their main target is not focused on reducing the overhead associated with the volume of data involved in the sampling process. This aspect directly impacts on monitoring costs and efficiency.

In this context, this paper presents a new multiadaptive traffic sampling technique, based on linear prediction, which aims to reduce the volume of data considered in the network measurements without compromising estimation accuracy. With this technique, the measurement overhead is significantly reduced, while assuring that sampled traffic reflects the statistical characteristics of the global traffic under analysis.

For this purpose, the traffic selection process considers the levels of network activity, and is configured to reduce measurement impact when the network activity increases and measurement processes tend to be heavier for measurement points. The multiadaptive behavior of this proposal is achieved considering both the sampling interval and sample size as adaptive parameters, bounded by thresholds that guarantee the representativeness of samples in capturing network behavior. A proof-of-concept is provided using real traffic traces representing distinct traffic scenarios.

The remaining of this document is organized as follows: a debate on representative sampling approaches and the motivation for the present proposal is included in Section II; the multiadaptive sampling technique and its design goals are described in Section III; the methodology for evaluating the performance of the technique in different traffic scenarios is presented in Section IV; the evaluation results are discussed in Section V; finally, the conclusions are drawn in Section VI.

978-1-4673-1544-9/12/$31.00 ©2012 IEEE

II. Related Work

Adaptive sampling measurement techniques are generally developed for a specific estimation parameter, such as packet loss [4] [5] or delay and jitter [6] [7]. A larger group of proposals are aimed at traffic characterization and SLA monitoring [8] [9] [10] [7].

Common generalist techniques are usually based on Fuzzy Logic and Linear Prediction. In adaptive sampling based on fuzzy logic [11] [12], a controller adjusts the sampling rate based on past similar experiences, determining the most appropriate action for current traffic conditions [13]. This approach requires a long-term database to store the knowledge and the possible action for each situation.

Linear prediction based techniques [14] [15] try to forecast network behavior based on an observed parameter in past samples. In these techniques, when the prediction is correct, the sampling rate can be reduced, while inaccurate predictions indicate a change in network activity and, therefore, an increase in sampling rate is required to determine the new pattern behavior [13]. In this sampling approach just a fixed number of samples is stored by the measurement point which are used in the prediction process. However, if the sampling frequency increases more resources will be required from the measurement point, precisely at the most critical moment of its operation.

The sampling technique proposed in this paper attempts to decrease resource consumption related to the processing, storage and transmission of captured packets in high network activity periods, while maintaining the accuracy of network statistical behavior estimation. The following section details the multiadaptive sampling technique design goals and operation.

III. Multiadaptive Sampling Technique

A. Concepts

Traffic sampling techniques share a set of concepts sometimes presented in an ambiguous way. To avoid inconsistencies, it is assumed the most common terms in accordance with the following definitions:

- *Sample:* subset of the network packets that are selected by the measurement point and are considered in the estimation of network parameters. These packets are also used by the adaptive measurement algorithms as input for the estimation of the reference parameter;
- *Sample size:* time interval in which all incoming packets at the measurement point are selected and captured to compose a sample unit;
- *Interval between samples:* time interval in which all incoming packets are ignored for measurement purposes. During this time, the behavior of the network is not considered for the parameters estimation and for the configuration of the adaptive sampling rate;
- *Reference parameter:* observed value in each sample, e.g. throughput, used as measurement parameter and input for configuring the sampling adaptation parameters.

Figure 1 illustrates the above concepts.



Fig. 1. Conceptual description

B. Design Goals

The multiadaptive sampling technique proposed in this work improves the sampling reactivity and efficiency, considering both the interval between samples and the sample size as adjustable parameters. In this way, the present proposal attends to the following design goals:

1) the adaptive nature of the technique should be driven by simplicity of implementation and low consumption of resources. This motivates the adoption of linear prediction sampling;
2) the sampling frequency must be increased whenever there is a noticeable increase in network activity to allow detecting conveniently the new traffic pattern. This is achieved through the reduction of the interval between samples;
3) as increasing the sampling frequency implies a higher consumption of resources (processing and storage), the sample size should be reduced to mitigate overhead while maintaining the representativeness of samples in the estimation of the parameters under analysis;
4) the adaptive parameters should be properly bounded by thresholds to guarantee representative data for capturing network behavior.

In [13] a linear prediction function was presented for adapting sampling frequency. While this technique allows to meet requirements (1) and (2) its processing overhead is still significant as the whole network traffic, regardless of whether or not the packets belong to a sample, is considered for the definition of the reference parameter.

In the present proposal the technique in [13] is adapted considering only packets belonging to previously collected samples to drive future sampling decisions, clearly reducing the processing overhead. The adapted predictive function is presented in Equation 1. The set of rules and calculations determining the change factor in the interval between samples are presented and discussed in Equation 2 and Table I.

As none of the techniques published so far present features for satisfying requirement (3), this paper proposes a set of rules for varying the sample size according to the level of network activity. These rules increases and decreases the sample size differently considering the observed network behavior. The dynamic variation in sample size complies with the criteria presented in Table II.

The thresholds to bound the adaptive parameters to meet requirement (4) are set experimentally and discussed throughout the following section.

C. Multiadaptive Sampling Technique Description

The multiadaptive technique takes into account the last N samples to estimate the future value of the reference parameter,

which is then used to determine the next interval between samples and the size of the next sample. Thus, for a sampler of order N, the expected value Xp of the reference parameter for the next collected sample is defined by Equation 1.

$$Xp = X[N] + \frac{\Delta T_{current}}{N-1} \sum_{i=1}^{N-1} \left(\left| \frac{X[i+1] - X[i]}{\Delta T[i]} \right| \right) \quad (1)$$

To estimate the value of Xp a vector X stores the values of the reference parameter of the last N samples, where $X[N]$ refers to the most recent sample. A second vector T stores the intervals between samples, where each ΔT is the time between the end of the sample $X[i]$ and the beginning of the sample $X[i+1]$. Therefore, $\Delta T_{current}$ is the interval between samples $X[N]$ and $X[N-1]$.

1) Defining the interval between samples: When a new sample is collected, the corresponding value of the reference parameter S is compared with the expected value Xp. Based on a factor of change m, obtained comparing $Xp - X[N]$ and $S - X[N]$, it is applied a set of rules to define ΔT_{next}, used to schedule the start of the next sample. The factor m is given by Equation 2.

$$m = \left| \frac{Xp - X[N]}{S - X[N]} \right| \quad (2)$$

Equation 2 returns a result close to 1 when the expected value Xp is close to the current value of S. The range of values that satisfies this condition is defined as:

$$m_{min} < 1 < m_{max}$$

If $m < m_{min}$ the reference parameter is changing faster than predicted. This behavior indicates more network activity than expected. Thus, the interval between samples should be decreased to achieve more accurate values, which will be used in following predictions.

On the other hand, if $m > m_{max}$ the value of the reference parameter is changing slower than predicted, then the interval between samples can be increased, reducing the measurement impact.

If the value of S does not change since the last sample $X[N]$, m assumes an undefined value, which indicates that the network is stable. In this case, the interval between samples is exponentially increased.

Table I lists the rules used to generate ΔT_{next} from the current value of m. The values of m_{max} and m_{min} are set experimentally as $m_{min} = 0.9$ and $m_{max} = 1.1$ due to its good performance for multiple traffic types [13].

An additional constraint is used to prevent ΔT from growing indefinitely, thus maintaining a minimum number of samples to ensure data to new predictions. Conversely, the maximum frequency of sampling is also limited, so that the sampling interval does not tend to zero, which would result in capturing all traffic. These limits were tuned experimentally and define the minimum interval between samples as 0.1s and the maximum interval between samples as 8s.

TABLE I
RULES TO DEFINE THE NEXT INTERVAL BETWEEN SAMPLES [13]

current m	ΔT_{next}
$m < m_{min}$	$\Delta T_{next} = m * \Delta T_{current}$
$m_{min} \leq m \leq m_{max}$	$\Delta T_{next} = \Delta T_{current}$
$m_{max} < m$	$\Delta T_{next} = \Delta T_{current} + 1seg$
$m_{undefined}$	$\Delta T_{next} = 2 * \Delta T_{current}$

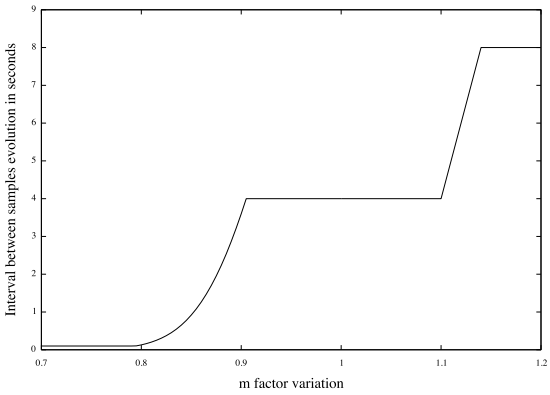


Fig. 2. Evolution of ΔT

Figure 2 illustrates the evolution of the interval between samples as a function of a linear variation of the m factor. As shown, the reactivity is higher when the network activity increases, i.e., when $m < m_{min}$. If the order of magnitude in the factor variation is above 0.01, there will be a significant change in the curve. The same does not occur for low network activity, i.e., when $m > m_{max}$. In this case, the interval between samples varies linearly.

2) Defining the sample size: For adapting the sample size it is also considered the indication of network activity provided by m factor. However, these changes are linear with order equal to $1/k$. This ensures that a sudden change in network behavior does not produce very different sample sizes.

The parameter k was firstly presented and experimentally set in [7], aiming at defining the change in the interval between samples based on a set of comparative statistics between two adjacent samples. In the present technique, the parameter was adopted as the changing factor in the sample size based on a number N of samples previously collected.

Table II presents the rules used to define the next sample size, where $\Delta S_{current}$ represents the current sample size and ΔS_{next} the size of the next sample to be collected.

TABLE II
RULES TO DEFINE THE NEXT SAMPLE SIZE

Current sample size	Next sample size
$m < m_{min}$	$\Delta S_{next} = \Delta S_{current} - (\Delta S_{current}/k)$, $k = 4$
$m_{min} \leq m \leq m_{max}$	$\Delta S_{next} = \Delta S_{current}$
$m_{max} < m$	$\Delta S_{next} = \Delta S_{current} + (\Delta S_{current}/k)$, $k = 10$
$m_{undefined}$	$\Delta S_{next} = \Delta S_{current} + (\Delta S_{current}/k)$, $k = 10$

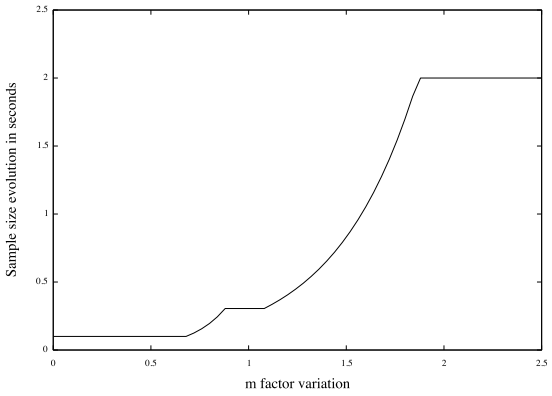


Fig. 3. Sample size evolution

According to Table II, in moments of less activity in the network, the sample size is increased by a factor of $1/k$ with $k = 10$. This allows to collect more information about the network in less critical periods of its operation, in sparser sampling events. In moments of increased activity, the sample size is decreased by a factor of $1/k$ with $k = 4$. This reduction in sample size, associated with the higher frequency in the sampling process, intents to reduce the measuring overhead in the measurement point.

Similarly to the definition of the interval between samples, restrictions were defined to limit sample sizes variation. These restrictions avoid samples extremely small, which difficult estimating parameters statistically, as well as samples excessively large, closely matching a capture of the total traffic. These limits were set experimentally, determining the minimum sample size as 0.1s and the maximum sample size as 2s.

Figure 3 illustrates the evolution of the sample size as a function of a linear variation of the m factor. As shown, similarly to the interval between samples, the variation in sample size is more reactive when the network activity increases. However, in this case, for both low and high observed activity, changes in sample size are constant, following the parameter k in each situation.

IV. Methodology of tests

The proof-of-concept of the multiadaptive sampling strategy (MA) aims to determine its ability to correctly capture the network behavior with reduced overhead. Therefore, a statistical analysis of the reference parameter obtained through traffic sampling and using the total traffic is carried out. In addition, the performance of the proposed strategy is compared to other commonly used techniques, the Systematic Time-based technique (ST) [2] and the Adaptive Linear Prediction technique (LP) [13]. For this purpose a set of statistical parameters is evaluated for four different traffic types, aiming at demonstrating the effectiveness of the proposal in distinct traffic scenarios.

The reference parameter adopted in the tests is network throughput, which is a proper indicator about the network activity variation. The order of prediction N is set as three.

A. Traffic scenarios

The performance of the multiadaptive technique was assessed using real traffic traces representing distinct traffic scenarios. Note that, although evaluation tests are based on traffic traces previously captured, the adaptive parameters of the sampling techniques are configured dynamically during the sampling process. Similarly to a real-time operational environment, there are no previous knowledge of the characteristics of the traffic passing through.

1) ***SIGCOMM 08****:* This trace, captured during approximately eight hours of the SIGCOMM 2008 conference, includes all the conference participants communications through IEEE 802.11a access points [16].

2) ***OC-48****:* This trace corresponds to a capture of approximately 5 minutes in a backbone link OC - Optical Carrier 48, which has a nominal throughput of 2.48 Gbps. The capture was carried out, passively, in a large ISP from the US West Coast and is available at CAIDA - The Cooperative Association for Internet Data Analysis [17].

3) ***VoIP-SIP****:* This trace represents around twenty hours of VoIP traffic, using SIP and G.711 encoding, in a Brazilian university campus with Fast Ethernet links.

4) ***Video Streaming****:* This trace corresponds to the capture of a live stream transmission of the STS-135 launching mission conducted by NASA on July 8, 2011. The stream was broadcast in High Definition 720p, through NASA TV channel and encoded in MPEG-4 using RTP as transport protocol.

B. Benchmark Techniques

1) Systematic Time-based Sampling (ST): This standard technique [2] is one of the most common sampling technique currently used. In the sampling process the packet selection follows a deterministic function based on arrival time to the measurement point, i.e., the sample size and the time between samples are set at the beginning and are unchanged along the sampling process. In this work, as suggested in [13], the operational parameters have been set to sample size equals to 0.1s and time between samples equals to 0.5s.

2) Adaptive Linear Prediction Sampling (LP): This adaptive sampling technique was proposed by [13]. This technique adjusts the interval between samples based on the level of network activity using linear prediction. However the sample size is fixed along the sampling process.

Here, the initial interval between samples was set to 0.2s and the sample size was fixed as 0.1s. Another configurable parameter of this technique is the order of prediction, which has been set as defined in the original specification [13] and corresponds to 2. Although the technique takes into account the last two samples to decide the next interval between samples, the measurement point still needs to maintain information about all traffic since the beginning of its execution. This is needed because the technique analyzes the evolution of the reference parameter based on the accumulated amount of data until the end of the sample N relative to the sample $N - 1$.

C. Statistical Parameters

The comparative statistical parameters aim at determining the equivalence between the total traffic behavior and the behavior resulting from the sampling process. Additionally, they intent to quantify the efficiency of the strategy facing other proposals.

Although the statistical parameters are common in most representative research works on adaptive sampling, this work extends previous works by cross-checking results from sampling and total traffic. The common approach used in evaluation tests only compares the performance of the proposed techniques to the systematic technique.

The parameters used on tests are described below.

1) ***Overhead****:* The main goal of the multiadaptive technique is to maintain the representativeness of the sampled traffic, while reducing the overhead associated with the amount of packets processed, stored and transmitted by the measurement point. This goal is verified by comparing techniques using the following parameters:

- *Number of Packets:* corresponds to the total number of packets captured during the sampling process for each sampling technique;

- *Amount of Data:* the sum, in Mbyte, of all packets collected with each sampling technique. For this metric, it is used the total length field within IP header;

- *Number of Samples:* total number of samples captured during the sampling process. This parameter is relevant because, for each new sample, resources are required to evaluate the sampling adaptation parameters, namely the new sample size and the interval until it starts. Thus, the higher the number of samples is, the more calculations at the measurement point are required.

2) ***Throughput****:* The most common way to analyze the traffic behavior is through the instantaneous throughput, measuring the amount of data traffic in a time interval, usually in kbps. This metric is commonly used for a graphical and statistical representation of the network status in real time.

Related parameters are the explosivity and the coefficient of variation (CV). These metrics belong to descriptive statistics and measure the variability of time series. In network measurements, they help to identify and characterize the burstiness of traffic. Explosivity is obtained by the ratio between the peak and average throughput in a measurement interval. The coefficient of variation is obtained through the ratio between standard deviation and average throughput.

The correlation between the sampled traffic and total traffic is also evaluated, studying statistically the relationship between the corresponding series. The coefficient of correlation ρ is obtained by the Pearson method [18], where:

$$0.7 < \rho \leq 1, \text{ indicates a strong correlation}$$
$$0.3 < \rho \leq 0.7, \text{ indicates a moderate correlation}$$
$$0 \leq \rho \leq 0.3, \text{ indicates a weak correlation}$$

To assess the discrepancy between the mean of the total traffic and its sampled version [7], the relative mean error is calculated using:

$$RME = \frac{|M_{total} - M_{estimated}|}{M_{total}} \quad (3)$$

where M_{total} is the average throughput of total traffic and $M_{estimated}$ is the average throughput of the sampled traffic.

V. Evaluation Results

This section presents and discusses the results achieved from the evaluation of the multiadaptive sampling technique, using the four traffic types and statistical parameters defined above. As mentioned, Systematic and LP techniques are used for comparison purposes.

A. Overhead reduction

Benchmarking carried out show that the multiadaptive technique leads to a significant overhead reduction for all traffic types under consideration. As an example, Figure 4 illustrates the number of packets, the data volume and the number of samples for OC-48 traffic, when applying each sampling technique and facing the total traffic. Note that, in this case, the sampled traffic resulting from the multiadaptive technique corresponds to 5.14% of total data in the original trace. Table III extends the results above, quantifying the overhead for all traffic types.

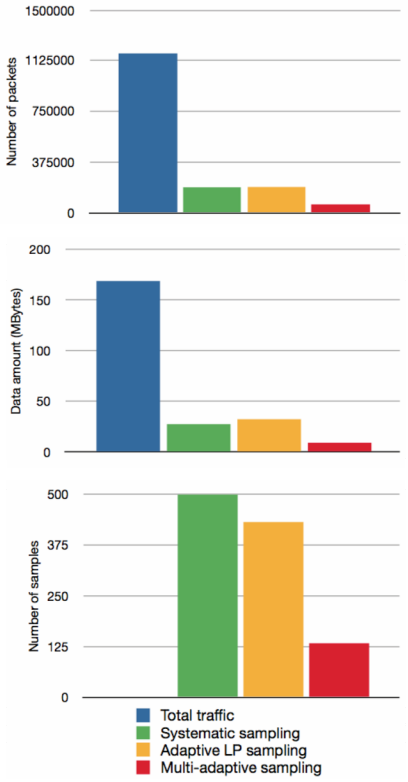


Fig. 4. Overhead reduction: OC-48

Comparing to the systematic technique, the multiadaptive technique reduces in 68.87% the number of collected packets, in 68.94% the total amount of data sampled, and in 73.40% the number of samples used to characterize the OC-48 traffic. For

TABLE III
OVERHEAD REDUCTION FOR ALL TRAFFIC TYPES

Traffic / Parameter	Total	ST	LP	MA
OC-48				
Number of packets	6550395	1094002	935066	340577
Data volume (MBytes)	3189.67	533.39	454.92	165.72
Number of samples		500	431	133
SIP VoIP				
Number of packets	1172378	188471	190572	62650
Data volume (MBytes)	168.68	27.41	32.27	9.03
Number of samples		105385	53081	27176
SIGCOMM08				
Number of packets	4513615	749977	899145	226060
Data volume (MBytes)	2382.92	395.97	511.93	120.49
Number of samples		44511	37104	9598
Video streaming				
Number of packets	97248	16303	13705	4711
Data volume (MBytes)	101.76	17.02	14.86	4.85
Number of samples		1090	406	362

the same traffic type, regarding the adaptive LP technique, the multiadaptive technique achieves a reduction of 63.58% in the number of collected packets, 63.58% in the total data amount, and 69.14% in the number of processed samples. As shown in Table III, for the remaining traffic types the overhead reduction achieved is similar.

These results clearly attest the performance improvement for applying the proposed technique in measurement points. However, to prove measurement efficiency the statistical representativeness of the sampled traffic has to be evaluated, verifying its ability to capture real traffic behavior. This evaluation is provided below.

B. Throughput estimation

Table IV presents the statistical benchmarking resulting from applying each sampling technique for throughput estimation in all traffic traces. The results show that systematic and multiadaptive techniques achieve high-quality representations of traffic behavior when comparing average throughput, explosivity, CV and correlation with the corresponding total traffic statistics. The adaptive LP technique tends to overestimate throughput when using low data rate traffic traces, particularly visible in VoIP traffic. The close match of this technique was achieved for OC-48 trace, which represents a high throughput link of a ISP. The correlation analysis of the instantaneous throughput time series (considering a measurement interval of 1s), obtained comparing each sampling strategy to the original trace, presents high coefficients of correlation.

Regarding the multiadaptive technique, when analysing the trade-off between accuracy and overhead reduction, see Tables IV and III, it is clear that this technique promotes a significant improvement to the other sampling techniques considered.

The accuracy achieved by the multiadaptive technique is ratified when considering the relative mean error of the estimated average throughput. Table V presents the relative mean error resulting from throughput estimation for each sampling technique applied to all traces.

TABLE IV
OVERALL ESTIMATED BEHAVIOR

Traffic / Parameter	Total	ST	LP	MA
SIGCOMM08				
Throughput	730.95	728.76	1130.27	803.65
Explosivity	14.39	15.16	10.05	12.93
C.V.	1.90	2.07	1.56	1.93
Correlation		0.91	0.91	0.89
SIP VoIP				
Throughput	21.85	21.30	50.49	22.03
Explosivity	158.40	132.14	58.00	112.72
C.V.	5.28	5.41	2.90	5.08
Correlation		0.71	0.88	0.86
OC-48				
Throughput	87103.35	87390.26	86467.86	86825.99
Explosivity	1.40	1.35	1.34	1.23
C.V.	0.10	0.09	0.10	0.13
Correlation		0.84	0.81	0.82
Video streaming				
Throughput	1275.84	1279.17	2998.53	1425.09
Explosivity	19.13	19.98	8.41	17.02
C.V.	3.93	3.96	2.50	3.66
Correlation		0.99	0.98	0.98

TABLE V
RELATIVE MEAN ERROR

Traffic	ST	LP	MA
SIGCOMM08	0.002	0.54	0.09
SIP VoIP	0.02	1.31	0.008
Víideo streaming	0.002	1.35	0.11
OC-48	0.003	0.007	0.003

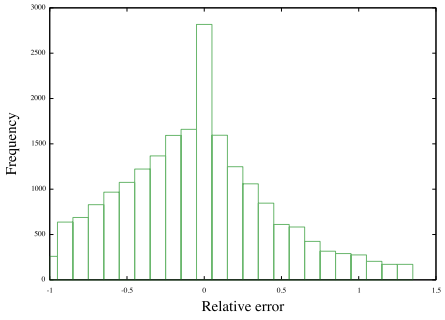
For a more detailed analysis, Figure 5 presents the relative error spread for each sampling technique when applied to the SIGCOMM trace. As shown, for all the tested techniques the error distribution is similar, concentrated near to zero and with a well-bounded tail. The figure also shows that the higher relative errors result from overestimation, expressed by the tail behavior in the positive values of the histogram.

From these results it is possible to verify that, despite the significant reduction on traffic volume considered, the multiadaptive technique ability to capture the real traffic behavior correctly is not compromised. For several statistical parameters, the proposed technique outperforms the conventional techniques which is even more significant when considering the high decrease on measurement overhead.

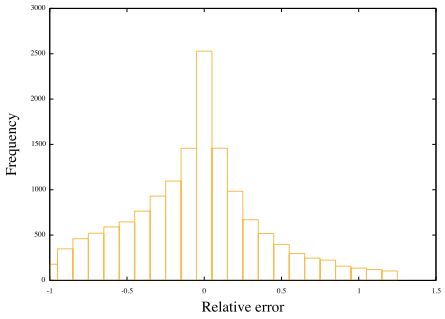
VI. CONCLUSIONS

This paper has presented a multiadaptive traffic sampling technique based on linear prediction. The multiadaptive functionality is associated with the ability to change the interval between samples and the sample size according to the network activity. Controlling these two variables and defining appropriate bounds for its values, this technique reduces significantly the overhead associated with the traffic sampling process while maintaining the accuracy in estimating the statistical network behavior when taking throughput as reference parameter.

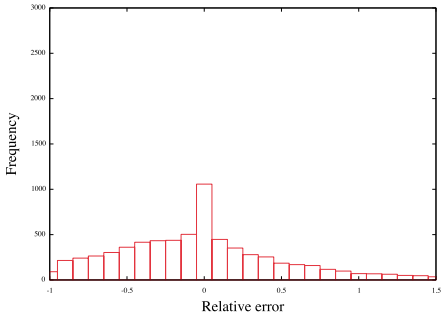
The performance evaluation results, using real traffic representing distinct profiles, have demonstrated the effectiveness of



(a) Systematic



(b) Adaptive LP



(c) Multiadaptive

Fig. 5. Error frequency: SIGCOMM08

the technique, outperforming classic techniques both in accuracy and data volumes processed. Future work will be centred in further developing the technique regarding simultaneous estimation of parameters such as packet loss, delay and jitter, both in real time and in large-scale testbeds.

ACKNOWLEDGMENT

This work is funded by FEDER through the Competitiveness Factors Operational Programme - COMPETE and Portuguese National Funds through FCT - Foundation for Science and Technology under the Project: FCOMP-01-FEDER-0124-022674

REFERENCES

[1] S. K. Thompson, *Sampling*, 2nd ed. Wiley-Interscience, 2002.

[2] T. Zseby, M. Molina, and N. Duffield, "Sampling and Filtering Techniques for IP Packet Selection," RFC 5475, Internet Engineering Task Force, Mar. 2009. [Online]. Available: http://datatracker.ietf.org/doc/rfc5475/

[3] C. Henke, C. Schmoll, and T. Zseby, "Empirical evaluation of hash functions for packetid generation in sampled multipoint measurements," in *Passive and Active Network Measurement*, ser. Lecture Notes in Computer Science, S. Moon, R. Teixeira, and S. Uhlig, Eds. Springer Berlin / Heidelberg, 2009, vol. 5448, pp. 197–206. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-00975-4_20

[4] R. Serral-Gracia, A. Cabellos-Aparicio, and J. Domingo-Pascual, "Packet loss estimation using distributed adaptive sampling," in *Network Operations and Management Symposium Workshops, 2008. NOMS Workshops 2008. IEEE*, Apr. 2008, pp. 124 –131.

[5] J. Sommers, P. Barford, N. Duffield, and A. Ron, "Improving accuracy in end-to-end packet loss measurement," in *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '05. New York, NY, USA: ACM, 2005, pp. 157–168. [Online]. Available: http://doi.acm.org/10.1145/1080091.1080111

[6] Z.-G. Hu, D.-L. Zhang, C.-P. Hou, and J.-S. Zhang, "Adaptive sampling algorithm of network round-trip time," *Journal of Computer Applications*, vol. 30, no. 2, pp. 319–322, feb 2010.

[7] A. Dogman, R. Saatchi, and S. Al-Khayatt, "An adaptive statistical sampling technique for computer network traffic," in *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, Jul. 2010, pp. 479 –483.

[8] T. Zseby, T. Hirsch, and B. Claise, "Packet sampling for flow accounting: Challenges and limitations," in *Passive and Active Network Measurement*, ser. Lecture Notes in Computer Science, M. Claypool and S. Uhlig, Eds. Springer Berlin / Heidelberg, 2008, vol. 4979, pp. 61–71. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-79232-1_7

[9] C. Hu, S. Wang, J. Tian, B. Liu, Y. Cheng, and Y. Chen, "Accurate and efficient traffic monitoring using adaptive non-linear sampling method," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, Apr. 2008, pp. 26 –30.

[10] A. N. Mahmood, J. Hu, Z. Tari, and C. Leckie, "Critical infrastructure protection: Resource efficient sampling to improve detection of less frequent patterns in network traffic," *Journal of Network and Computer Applications*, vol. 33, no. 4, pp. 491 – 502, 2010.

[11] J. Giertl, J. Baca, F. Jakab, and R. Andoga, "Adaptive sampling in measuring traffic parameters in a computer network using a fuzzy regulator and a neural network," *Cybernetics and Systems Analysis*, vol. 44, pp. 348–356, 2008, 10.1007/s10559-008-9005-0. [Online]. Available: http://dx.doi.org/10.1007/s10559-008-9005-0

[12] Q. Xin, L. Hong, and L. Fang, "A modified flc adaptive sampling method," in *Communications and Mobile Computing, 2009. CMC '09. WRI International Conference on*, vol. 2, Jan. 2009, pp. 515 –520.

[13] E. A. Hernandez, M. C. Chidester, and A. D. George, "Adaptive sampling for network management," *Journal of Network and Systems Management*, vol. 9, pp. 409–434, 2001, 10.1023/A:1012980307500. [Online]. Available: http://dx.doi.org/10.1023/A:1012980307500

[14] Y. Lu and C. He, "Resource allocation using adaptive linear prediction in wdm/tdm epons," *AEU - International Journal of Electronics and Communications*, vol. 64, no. 2, pp. 173 – 176, 2010.

[15] Y. Wei, J. Wang, and C. Wang, "A traffic prediction based bandwidth management algorithm of a future internet architecture," *Intelligent Networks and Intelligent Systems, International Workshop on*, vol. 0, pp. 560–563, 2010.

[16] A. Schulman, D. Levin, and N. Spring, "CRAWDAD trace umd/sigcomm2008/pcap/ethernet (v. 2009-03-02)," Downloaded from http://crawdad.cs.dartmouth.edu/umd/sigcomm2008/pcap/Ethernet, Mar. 2009.

[17] C. Shannon, E. Aben, kc claffy, D. Andersen, and N. Brownlee, "The CAIDA OC48 traces dataset - 20030424-005000-0-anon," Downloaded from http://www.caida.org/data/passive/passive_oc48_dataset.xm, Apr. 2003.

[18] K. Pearson, *On a New Method of Determining Correlation Between a Measured Character A, and a Character B, of which Only the Percentage of Cases Wherein B Exceeds (or Falls Short of) a Given Intensity is Recorded for Each Grade of A*, 1st ed. Biometrika, 1909, vol. 7, pp. 96–105.