Computer Networks 57 (2013) 3357–3369

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

A multiadaptive sampling technique for cost-effective network measurements

João Marco C. Silva, Paulo Carvalho, Solange Rito Lima *

Departamento de Informática, Centro Algoritmi, Universidade do Minho, Braga, Portugal

ARTICLE INFO

Article history:
Received 7 January 2013
Received in revised form 26 July 2013
Accepted 30 July 2013
Available online 8 August 2013

Keywords:
Sampling techniques
Traffic measurements
Linear prediction
Adaptive sampling

ABSTRACT

The deployment of efficient measurement solutions to assist network management tasks without interfering with normal network operation assumes a prominent role in today's high-speed networks attending to the huge amounts of traffic involved. From a myriad of proposals for traffic measurement, sampling techniques are particularly relevant contributing effectively for this purpose as only a subset of the overall traffic volume is handled for processing, preserving ideally the correct estimation of network statistical behavior.

In this context, this paper proposes MuST – a multiadaptive sampling technique based on linear prediction, aiming at reducing significantly the measurement overhead and still assuring that traffic samples reflect the statistical characteristics of the global network traffic under analysis. Conversely to current sampling techniques, MuST is a multi and self-adaptive technique as both the sample size and interval between samples are self-adjustable parameters according to the ongoing network activity and the accuracy of prediction achieved.

The tests carried out demonstrate that the proposed sampling technique is able to achieve accurate network estimations with reduced overhead, using throughput as reference parameter. The evaluation results, obtained resorting to real traffic traces representing wired and wireless aggregated traffic scenarios and actual network services, prove that the simplicity, flexibility and self-adaptability of the proposed technique can be successfully explored to improve network measurements efficiency over distinct traffic conditions. For optimization purposes, this paper also includes a study of the impact of varying the order of prediction, i.e., of considering different degrees of past memory in the self-adaptive estimation mechanism. The significance of the obtained results is demonstrated through statistical benchmarking.

1. Introduction

In today's Internet, network measurement techniques 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 an accurate view of the status of the network and of provided services. Therefore, key aspects to balance when designing an efficient network measurement solution are the estimation accuracy and the measurement overhead.

Traffic sampling techniques have been extensively used to reduce the impact of performing traffic measurements on operational networks. In these techniques, a subset of packets is selected and then used to estimate network parameters, avoiding processing all network traffic [1].

To face the drawbacks of common sampling techniques (deterministic or random), adaptive sampling techniques

* Corresponding author. Tel.: +351 253604436.
E-mail addresses: joaomarco@di.uminho.pt (J.M.C. Silva), pmc@di.uminho.pt (P. Carvalho), solange@di.uminho.pt (S.R. Lima).

1389-1286/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2013.07.023

have been proposed (see discussion in Section 2). In *Adaptive Sampling* techniques the packet selection process considers the value of a reference parameter (e.g., throughput) observed during a measurement period. In this way, the sampling process becomes more flexible and self-adaptive, i.e., the packet selection criterion may change dynamically over the measurement period.

Despite the evolution of adaptive techniques in estimating network performance parameters correctly, their main target has not been on reducing the overhead associated with data volume involved in the sampling process. This aspect directly impacts on monitoring costs and efficiency, being particularly limitative in high-speed networks handling massive traffic volumes. The efficiency of a sampling technique may be assessed through a proper balancing between estimation accuracy and measurement overhead.

In this context, this paper presents a new multiadaptive sampling technique (MuST) based on linear prediction, which aims to reduce the amount of data involved in the network measurements without compromising the estimation accuracy. Therefore, the main objective is to reduce the measurement overhead and still assure 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, being configured to reduce the measurement impact when the network activity increases or the measurement process tends to overload the measurement points (MPs). The multiadaptive behavior of the proposed technique is achieved considering both the interval between samples and the sample size as adaptive parameters, bounded by proper thresholds to guarantee the representativeness of samples in capturing the network behavior. A proof-of-concept is provided using real traffic traces representing distinct traffic scenarios. The results demonstrate the effectiveness and versatility of the present proposal, outperforming conventional sampling techniques.

The remaining of this document is organized as follows: the debate on representative sampling approaches and the motivation for the present proposal is included in Section 2; the multiadaptive sampling technique, its design goals, definition and operation are described in Section 3; the proof-of-concept objectives and methodology are presented in Section 4; the evaluation results are discussed in Section 5; and finally, the conclusions are drawn in Section 6.

2. Related work

Existing sampling techniques can 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 resulting from deterministic sampling may still overload measurement points and produce biased samples [2].

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 end-to-end 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].

Adaptive sampling 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 adaptive techniques are usually based on Fuzzy Logic or 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 the 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 then used in the prediction process. However, if the sampling frequency increases more resources will be required from the measurement point, precisely in a critical moment of its operation.

As discussed, current sampling techniques have as main objective estimating accurately parameters of interest regarding network status, but not necessarily the efficiency. The efficiency involves, beyond high accuracy, the ability to perform measurements with minor interference with the normal network operation. Therefore, the deployment of sampling techniques able to identify critical periods in network activity adjusting their dynamics accordingly, is crucial to reduce resource requirements on network nodes and to reach efficiency.

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

3. MuST – Multiadaptive Sampling Technique

3.1. Concepts

Traffic sampling techniques share a set of concepts sometimes presented in an ambiguous way. To avoid misunderstanding, the most common terms were adopted in accordance with the following definitions (see illustration in Fig. 1):

- *Sample:* subset of network packets that are selected at the measurement point and are considered in the estimation of network parameters. These packets are also