Enhancing Traffic Sampling Scope and Efficiency

João Marco C. Silva, Paulo Carvalho and Solange Rito Lima
Centro Algoritmi, Universidade do Minho, Braga, Portugal
Email: joaomarco@di.uminho.pt, pmc@di.uminho.pt, solange@di.uminho.pt

***Abstract*—Traffic Sampling is a crucial step towards scalable network measurements, enclosing manifold challenges. The wide variety of foreseeable sampling scenarios demands for a modular view of sampling components and features, grounded on a consistent architecture. Articulating the measurement scope, the required information model and the adequate sampling strategy is a major design issue for achieving an encompassing and efficient sampling solution. This is the main focus of the present work, where a layered architecture, a taxonomy of existing sampling techniques distinguishing their inner characteristics and a flexible framework able to combine these characteristics are introduced. In addition, a new multiadaptive technique proposal, based on linear prediction, allows to reduce the measurement overhead significantly, while assuring that traffic samples reflect the statistical behavior of the global traffic under analysis.**

I. Introduction

The massive traffic volumes and service heterogeneity in today's Internet urge for flexible, yet simple network measurement solutions to assist network and service management, without impairing network performance. Traffic sampling techniques assume a prominent role in this purpose as the amount of traffic processed can be substantively reduced, while keeping ideally a realistic view of network behavior.

Currently, traffic sampling sustains a wide range of network tasks (see Figure 1). For instance, its usefulness has been explored in: *traffic engineering* to assist traffic classification and characterization [1]; *network security* for anomaly and intrusion detection, botnet and DDoS identification [2]; *SLA compliance* and *QoS control* for estimating parameters such as delay, jitter and packet loss [3], [4]. Facing this multitude of contexts, a key aspect driving network sampling efforts should be centered on establishing a proper relation among the measurement scope, the required information model and the sampling strategy to adopt.

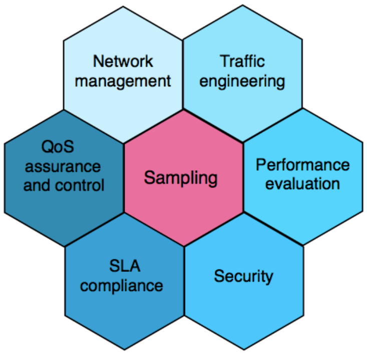

Fig. 1. Network tasks

Despite traffic sampling being a mandatory strategy for today and next generation networks, there are several challenges scarcely addressed in current and classical techniques, namely: (i) sampling techniques aim at estimating network performance parameters correctly, however their main target is not on optimizing the overhead associated with the data volume involved in the sampling process; (ii) the complexity and computational requirements of sampling algorithms are usually poorly analyzed, and it is still unclear which algorithm is the most suitable to deal with each traffic type [1]; and (iii) current sampling techniques only address a unique network monitoring activity. Therefore, a modular measurement solution based on traffic sampling able to support a wide range of network tasks efficiently is still an open issue.

In this context, the present work aims to devise an encompassing sampling-based measurement model for enhancing measurements in high-speed, large scale networks. The proposed model resorts to flexible traffic sampling techniques developed to improve the trade-off between measurement accuracy and overhead. Based on the work carried out so far, we present a three layer measurement architecture addressing the key components to sustain a versatile and lightweight measurement strategy. We propose a taxonomy of sampling techniques, driving the development of a traffic sampling framework, and a novel multiadaptive sampling technique which outperforms classic sampling techniques, both in accuracy and amount of data involved in the measurement process.

II. Measurement architecture

The sampling-based measurement architecture comprises three planes, as presented in Figure 2, and detailed below:

Management plane: This plane includes tasks directly deployed in measurement points or in external coordination entities. Based on specific requirements of each network task, measurement needs are identified and one or more measurement points are selected to be involved in the sampling process. This also involves identifying and selecting an information model able to define managed objects in the network independently of specific implementations or protocols in use, as recommended in RFC6727. The management plane is also responsible for processing measurement results reported by the control plane and providing a visualization component when applicable. The required processing may involve results of single or multipoint measurements.

Control plane: The control plane has a modular design, allowing a flexible sampling technique selection and configuration. Considering IETF PSAMP work and recent sampling proposals, a sampling taxonomy is here proposed to identify the inner characteristics distinguishing sampling techniques. This taxonomy also supports the definition of new sampling techniques able to be adjusted to each traffic/service measurement scenario. As shown, selecting and configuring a sampling technique involves electing the *granularity*, the *selection trigger* and the *selection scheme* to be applied, targeting a specific network task, accuracy requirements, computational weight and network activity. *Granularity* identifies the atomicity of the element under analysis in the sampling process: in a flow-level approach, flow attributes are used to guide sampling decisions; in a packet-level approach, packets are eligible as single independent entities; *Selection trigger*, used to decide the spatial and temporal sample boundaries, may use a time-based approach, a count-based approach or an event-based approach; *Selection scheme* identifies the function defining which traffic packets will be selected and collected; this scheme may follow a deterministic, a random or an adaptive function.

In the control plane, the sampled data received from the data plane is processed and relevant metrics are estimated according to the network task measurement needs. These metrics are then aggregated (both in time and space) and exported following IETF guidelines (RFC6728, RFC6313), and using IPFIX specifications.

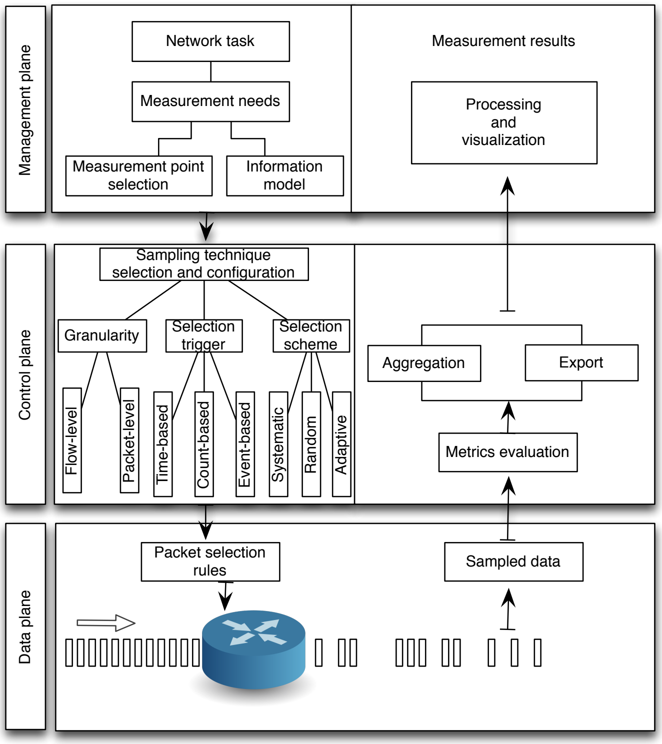

Fig. 2. Architecture description

Data plane: At data plane, traffic is collected from network interfaces by applying the sample rules defined in the control plane. Sampling typically resorts to tools available in network nodes, as systematic or random count-based sampling (e.g. NetFlow). The use of a flexible sampling framework will allow to configure the sampling technique according to the measurement purpose.

III. Ongoing Developments and Achievements

According to the measurement architecture and sampling taxonomy, a framework and a new multiadaptive sampling technique have been developed. The framework, implemented in Java using Jpcap, allows a flexible combination of the sampling features defined above, and may be applied to both online and offline measurement scenarios. The multiadaptive technique, based on linear prediction, considers the levels of network activity and is configured to reduce the measurement impact when the network activity increases or the measurement process tends to overload the measurement points. The multiadaptive behavior is achieved considering both the sampling interval and the sample size as adaptive parameters, bounded by proper thresholds to guarantee the representativeness of samples in capturing the network behavior. Using real traffic traces (OC48 from a US West Coast ISP, SIGCOMM08, SIP VoIP and Video streaming) the proposed technique is compared to conventional sampling techniques. As illustrated in Figure 3, multiadaptive sampling brings clearly an added value regarding the reduction of sampling data involved in the measurement processes without compromising the relative error in throughput estimation (for a complete description and statistical analysis see [5]). Future work will be centered on defining appropriate information models and sampling profiles adjusted to distinct network measurement needs.

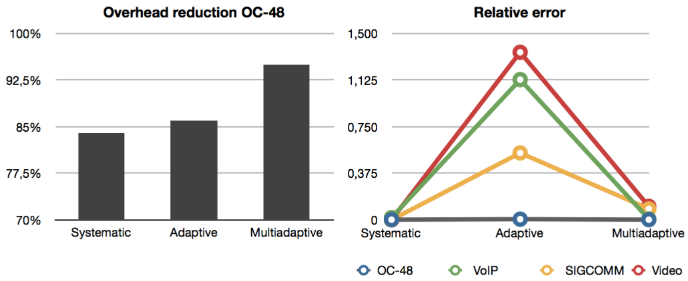

Fig. 3. Data overhead reduction and throughput estimation error

Acknowledgments - 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] D. Tammaro, S. Valenti, D. Rossi, and A. Pescapè, "Exploiting packet-sampling measurements for traffic characterization and classification," *International Journal of Network Management*, pp. 451–476, 2012.

[2] G. Androulidakis, V. Chatzigiannakis, and S. Papavassiliou, "Network anomaly detection and classification via opportunistic sampling," *Network, IEEE*, vol. 23, no. 1, pp. 6 –12, Jan.-Feb. 2009.

[3] J. Sommers, P. Barford, N. Duffield, and A. Ron, "Improving accuracy in end-to-end packet loss measurement," in *ACM SIGCOMM '05*. New York, NY, USA: ACM, 2005, pp. 157–168.

[4] 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 *IEEE INFOCOM 2008*, Apr. 2008, pp. 26 –30.

[5] J. M. C. Silva and S. R. Lima, "Multiadaptive sampling for lightweight network measurements," in *Computer Communications and Networks (ICCCN 2012)*, Aug. 2012, pp. 1–7.