Metherxis: Virtualized Network Functions for Micro-second Grade Latency Measurements

Diego R. Mafioletti,
Alextian B. Liberatto
IFES - Colatina/ES
loxxxa@gmail.com,
alextian@ifes.edu.br

Rodolfo S. Villaça,
Cristina K. Dominicini
IFES - Serra/ES
rodolfo.villaca@ifes.edu.br,
cristina.dominicini@ifes.edu.br

Magnos Martinello,
Moises R. N. Ribeiro
UFES - Vitória/ES
magnos@inf.ufes.br,
moises@ele.ufes.br

ABSTRACT

Network latency is critical to the success of many high-speed, low-latency applications. RFC 2544 discusses and defines a set of tests that can be used to describe the performance characteristics of a network device. However, most of the available measurement tools cannot perform all the tests as described in this standard. As a novel approach, this paper proposes Metherxis, a system that can be implemented on general purpose hardware and enables Virtualized Network Functions (VNFs) to measure network device latency with micro-second grade accuracy. Results show that Metherxis achieves highly accurate latency measurements when compared to OFLOPS, a well known measurement tool.

CCS Concepts

•**Networks → Network measurement;**

1. INTRODUCTION

Network latency measurements are crucial in providing reliable and efficient networked services, such as e-commerce, multimedia streaming, and social networking. Most of these service providers run its servers on clouds, which are geographically distributed around the world and far away from its users. Furthermore, highly accurate latency measurements are increasingly critical to the success of many high-speed, low-latency applications, such as trading transactions, databases that require improved timestamp accuracy, and server synchronization for automation or regulatory purposes.

A latency measurement is composed of 3 main components: propagation, transmission and processing delay. The first two components depend only on the distance, the physical media and the network bandwidth. Processing delay, in turn, depends on the computing system, which can be a server, a desktop, a smartphone or a network device. In terms of a network device latency measurement, the reference standard is the RFC 2544 [1], which discusses and defines a set of tests that can be used to describe its performance characteristics.

In the literature, there are several tools to evaluate the performance of network devices; among them, we highlight `pktgen` [3] and `OFLOPS` [2]. All the related tools suffer from at least one of the following limitations: do not provide the level of accuracy, flexibility and scalability required by high performance applications; cannot perform all the tests described in the RFC 2544 due to software or hardware limitations; or require high investment on specialized hardware.

To address these challenges, Metherxis is presented as a novel approach to measure latency with micro-second grade accuracy as a Virtualized Network Function (VNFs). Metherxis can be deployed on standard hardware and hosted in a Cloud or Datacenter infrastructure. The virtualization technique implemented by LXC allows Metherxis to be vertically scalable according to the number of network interfaces in the physical host and to deploy containers that share the same clock in order to measure the latency with high accuracy.

In Section 3, a loopback mode procedure is performed to compare Metherxis and `OFLOPS` tools. Moreover, this section evaluates the behavior of the selected tools under packet size and packet rate variations. Also, latency measurements are shown as the packet rate increases. Section 4 concludes the paper and points to future work.

2. METHERXIS

The key idea of Metherxis is to employ a single Linux host to allow the creation of a wide range of Virtualized Network Measurement Functions (VNMFs). To accomplish this task, it relies on LXC for resource isolation.

A VNMF in Metherxis is composed by two building blocks: a packet generator and a packet analyzer.

LANCOMM, August 22-26, 2016, Florianopolis , Brazil
© 2016 ACM. ISBN 978-1-4503-4426-5/16/08. . . $15.00
DOI: http://dx.doi.org/10.1145/2940116.2940131

The packet generator represents a sender (TX) and the packet analyzer represents a receiver (RX). Depending on the required measurement setup, one VNMF may require multiple senders (TX) and/or receivers (RX).

As shown in Fig 1, each building block has its own physical network interface and it is assigned to a specific namespace container. In this way, the building blocks can be vertically scalable from 1 to n. For measuring the latency with high accuracy, a packet generator can be deployed in namespace 0 (TX) generating traffic at physical ethernet port 0 and a packet analyzer can be set to receive packets in a namespace 1 (RX) at port 1.

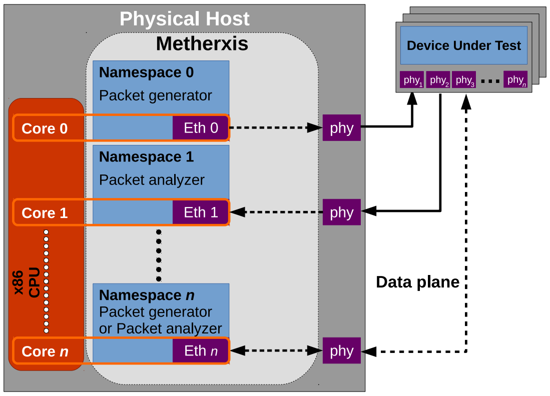

Figure 1: Metherxis conceptual view.

Metherxis Implementation: We chose pktgen [1] to generate packets in kernel mode. The latency measurement using pktgen requires the addition of a timestamp to the packet at its departure time. When the packet arrives at the receiver side, the receiver subtracts the current system clock from the marked timestamp.

However, pktgen is designed for a single host and is not optimized for the LXC concept. Thus, in order to enable the communication between the containers, Metherxis modifies pktgen and implements, in kernel mode, one timestamp array for each sender, which can be read by any receiver. Each array is uniquely identified by the sender source IP address and stores one tuple $< key, value >$ for each packet, which contains the packet identification and its timestamp, respectively. It is inserted in the Identification field of the IP protocol.

The timestamp is set into the array when the packet leaves the kernel to the network interface. At the receiver (RX), Metherxis reads the system clock at the packet arrival time subtracting it by the timestamp value stored at the position key of the array. In contrast to pktgen, that adds timestamps into the packet, Metherxis does not modify the payload, reducing the packet processing time and, consequently, increasing the measurement accuracy. The implementation is scalable since only the sender writes at its own array and any receiver is able to read timestamp values from any array.

3. BENCHMARKING METHERXIS

This section aims at evaluating the accuracy tradeoffs provided by different measurement approaches: OFLOPS x Metherxis. In order to define a widely known set of tests, our benchmarks follow the RFC 2544 [1]. Each experiment was repeated 30 times and we plotted the average results with a 95% confidence interval. For the loopback tests, we created a physical loopback by connecting the sender port to the receiver port.

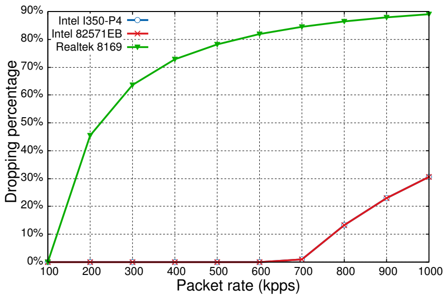

Figure 2: Dropping Packets Evaluation of Network Interface Cards

As the first step, we evaluated three 1 Gbps NIC models by performing the tests defined in the RFC 2544 with packet size of 64 Bytes. The goal of this test was to evaluate the NIC bottlenecks. Figs. 2 and 3 compares the number of dropped packets and latency values when varying packet rates from 100 to 1000 kpps. Beyond 1000 kpps all tested NICs are no longer loss free.

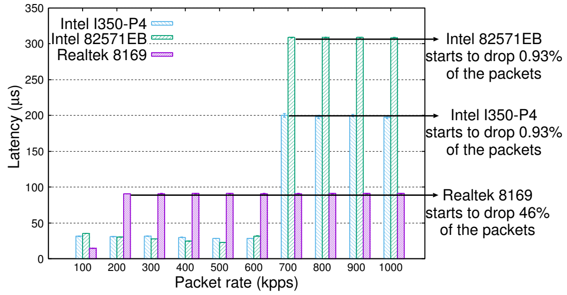

Figure 3: Latency Evaluation of Network Interface Cards

In Fig 2 both Intel NICs reached the same results, but Fig. 3 presents latency assessment at micro-second grade for the successfully transmitted packets. Note that latency tends to be insensitive to packet rate up to the point when packet losses start to happen. After that, a higher latency plateau is reached with severe packet loss. Intel I350-P4 was the last NIC to reach the second plateau. Thus, this NIC was selected to the remaining evaluations.

The next benchmark is a comparative with OFLOPS[2]. The goal is to evaluate the percentage of packets sampled (time-stamped) for latency calculation in both tools.

Fig. 4 shows a comparison between OFLOPS and Metherxis, both in loopback mode. It is clear that OFLOPS approach quickly limits measurement statis-

[1] https://pktgen.readthedocs.org/en/latest/

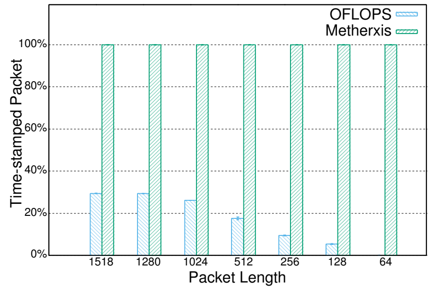

Figure 4: Percentage of time-stamped packets for variable packet size in loopback

tics by severely reducing the percentage of selected packets as packet lengths decrease. Even with the standard MTU size, less than 30% of transmitted packets are used for latency calculations. OFLOPS limitation is critical at 64 Bytes packets as no packets are used for latency computation. As a result, OFLOPS cannot be used to RFC 2544 complaint tests. On the other hand, the lightweight system adopted by Metherxis at kernel level not only generates and receives packets at wire speed, but it also considers 100% of the packets.

In order to further investigate OFLOPS' limitations, another test was performed by setting packet length at the shortest RFC 2544 packet size that OFLOPS can support (128 Bytes). Then, we selected different packet rates in order to evaluate its effect on the percentage of time-stamped packets. As seen in Fig. 5, in OFLOPS, latency is computed with a very few useful samples as the packet rate increases.

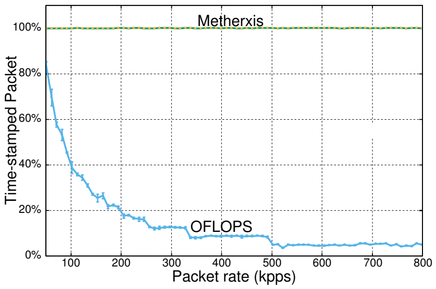

Figure 5: Percentage of time-stamped packets for variable packet rate in loopback

Fig. 5 clearly shows how packet rate affects OFLOPS framework as it barely reach 20% time-stamped packets at 200 kpps. In contrast, Metherxis is able to support 100% time-stamping with different packet rates.

More importantly, however, are the latency measurements in loopback mode, as they set the tool baseline accuracy. To this end, Fig. 6 presents OFLOPS and Metherxis latency measurements for 128 Bytes packet size, as packet rates increase. As expected, OFLOPS presents virtually no latency in loopback for packet rates around 500 kpps. Nevertheless, for packet rates above 500 kpps, its loopback latency jumps to a 38 μs plateau. On the other hand, Metherxis can reach stable and well-behaved loopback latency at different packet rates. Using off-the-shelf NICs, Metherxis reaches between 32 and 36 μs for 200 kpps and beyond.

It is worth mentioning that, despite its variation for lower packet rates, the 95% confidence interval bar is not visible for the whole range. This fact allows us to calibrate Metherxis for any packet rate, since the average values are reliable references that can be later subtracted from the actual measurements taken from the DUT (Device under Test). As a result, Metherxis enables inexpensive micro-second grade measurement tools to be built using its basic blocks.

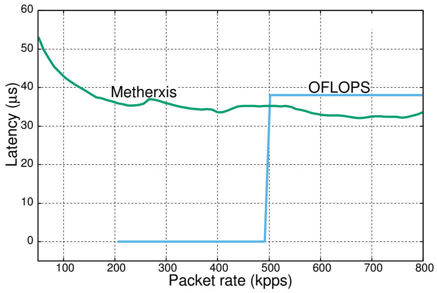

Figure 6: Metherxis x OFLOPS latency measurements for 128 Bytes packet and variable packet rate

4. CONCLUSION AND FUTURE WORK

This paper presented a new concept for multi-port micro-second grade latency measurements from inexpensive off-the-shelf x86 hardware. This was only possible thanks to the new concept of LXC. Physical loopback measurements were used to benchmark the system in comparison to OFLOPS. This opens new avenues for latency measurements and scalable VMNF, whereas complex measurements can be created using Metherxis.

5. REFERENCES

[1] Bradner, S., and McQuaid, J. Rfc2544: Benchmarking methodology for network interconnect devices. Website, 1999. http://www.ietf.org/rfc/rfc2544.txt.

[2] Rotsos, C., Sarrar, N., Uhlig, S., Sherwood, R., and Moore, A. W. Oflops: An open framework for openflow switch evaluation. In *13th International Conference on Passive and Active Measurement* (Berlin, Heidelberg, 2012), PAM'12, Springer-Verlag, pp. 85–95.

[3] Torrents, D. T. Open Source Traffic Analyzer. Master's thesis, KTH Information and Communication Technology, Stockholm, Sweden, 8 2010.

Acknowledgments

This work has received funding from CNPq, CAPES and FAPES. In addition, it is part of the FUTEBOL project, which has received funding from the European Union's Horizon 2020 for research, technological development, and demonstration under grant agreement no. 688941 (FUTEBOL), as well from the Brazilian Ministry of Science, Technology and Innovation (MCTI) through RNP and CTIC.