Automated Traffic measurements and analysis in Grid5000

Patrick Loiseau, Romaric Guillier, Oana Goga, Matthieu Imbert, Paulo Goncalves,
Pascale Vicat-Blanc Primet[*]
INRIA, ENS-Lyon, University of Lyon, France
{FirstName.Name}@ens-lyon.fr

Context and objectives

Today, the capacity and the usage of the Internet is fundamentally changing. In the forthcoming years, millions of homes will have access to the Internet with fibber lines. Network companies will offer speeds of up to 1 Gbps on fibbers. Consequently ultra-high-speed applications will be enabled for low-cost, such as high-definition teleconferencing, telemedicine and advanced telecommuting for people working from home. For understanding the Future Internet traffic and the required protocol's characteristics and behavior, the analysis of fully controlled and reproducible experiments which allow the experimenter to change a lot of parameters (aggregation level, congestion intensity, packet size or source flow size distribution, etc.); and then study separately their impact on the traffic are needed. To have a better insight on the real nature and on the evolution of network traffic we argue that fine-grain analysis of real traffic traces have to complement simulations studies as well as coarse grain measurement performed by classical flow measurement systems. In particular, packet level measurements and analysis are needed.

As will be demonstrated the combination of three elements: *Metroflux*, a very fine grain traffic capture and analysis tool, *Grid5000*, a very flexible large scale testbed and *NXE*, a network experiment automator, give researchers and network operators a unique and very powerful platform to perform packet level analysis of the traffic from a large collection of controlled sources, and particularly to perform the challenging task of monitoring a router buffer. We will illustrate several usages of this toolset, such as the investigation of conditions under which several traffic theories apply, as well as studies on traffic, protocols and systems interactions, etc.

Description of the experimental platform

Metroflux. *Metroflux* is a programmable system for packet and flow analysis which currently operates on 1 Gbps and 10 Gbps bi-directional links without loss. This system is composed of hardware elements and software components. It integrates the GtrcNet-1 box or the GtrcNet-10 box (for 10 Gbps links, see Fig. 1), a FPGA-based device technology for the capture of packet headers; and a storage server with large amount of disk space. The system is able to capture the first 52 bytes (for GtrcNet-1) or 56 bytes (for GtrcNet-10) of every packets, and to add a 60ns resolution timestamp to each of them. The duration of a capture session depends on the packet size and the throughput of the link, and can encompass several hours. The *Metroflux* system[1] also integrates original statistical analysis tools.

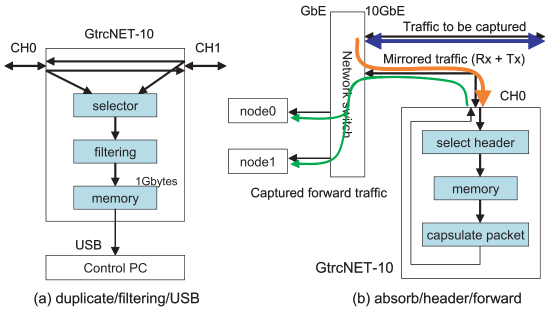

Figure 1: Example of packet capture on GtrcNET-10

Grid5000. *Grid5000* is a research tool, featured with deep control, reconfiguration and monitoring capabilities to complement network simulators and emulators. It allows the users to reserve the same set of dedicated nodes across successive experiments, and to have full controll of these nodes to run their own experimental condition injector and measurement software. *Grid5000* is a 5000 CPUs nation wide grid infrastructure dedicated to network and grid computing research. It involves 9 geographically distributed sites hosting one or more cluster of about 500 cores each (Fig. 2). The sites are interconnected by a dedicated optical network provided by RENATER, the French National Research and Education Network. It is composed of private 10 Gbps

[*]This work has been funded by the French ministry of Education and Research, INRIA, and CNRS, via ACI GRID's Grid'5000 project , ANR HIPCAL grant, the ANR IGTMD grant, IST EC-GIN project, INRIA GridNet-FJ grant, NEGST CNRS-JSP project.

[1]http://www.ens-lyon.fr/LIP/RESO/ Software/MetroFlux/index.html

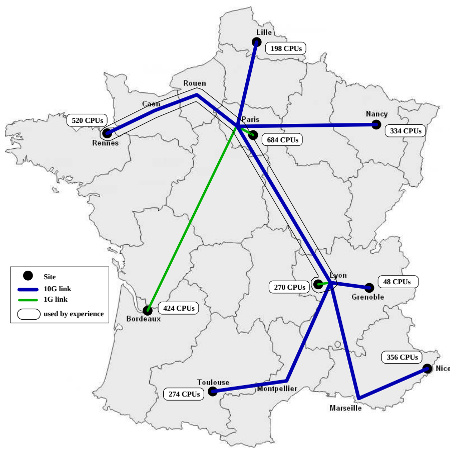

Figure 2: *Grid5000*

Ethernet links connected to a DWDM core with dedicated 10 Gbps lambdas. Two international interconnections are also available: one at 10 Gbps with DAS3 (the Netherlands) and one at 1 Gbps with Naregi (Japan). As a private testbed dedicated to research, *Grid5000* made it easy to install the experimental hardware of *Metroflux* at representative traffic aggregation points.

NXE. To handle the execution of large scale controlled experiments, we have developed a tool, *NXE*, that fully automates each step of the experiment execution on a real testbed from the topology definition to the log gathering, traffic collecting. The Network eXperiment Engine (*NXE*) is a tool to automate networking experiments in real testbeds environments. It allows to simply script experiments involving hundreds of nodes. More detailed information can be found on its software description page[2].

Experiments

Goals. The goal of this experiment is to investigate the specific impacts different traffic load features have on a bottleneck queue response. In particular, we propose to characterize the effect of sources dependence on the queuing delay time series. Using *Grid5000* and *NXE*, it is easy to steer a large number of sources with a controlled and versatile nature of interdependence: arrival times, durations, instantaneous throughputs, etc. As for *Metroflux* system, it allows for synchronously monitoring the input and the output of a router buffer at the packet level.

Scenarios. The topology used for this experiment is described on Figure 3: A set of N controlled sources is emitting towards the input buffer of switch router. Entry 1 of *Metroflux* system monitors the traffic corresponding to one

[2]http://www.ens-lyon.fr/LIP/RESO/Software/NXE/index.html

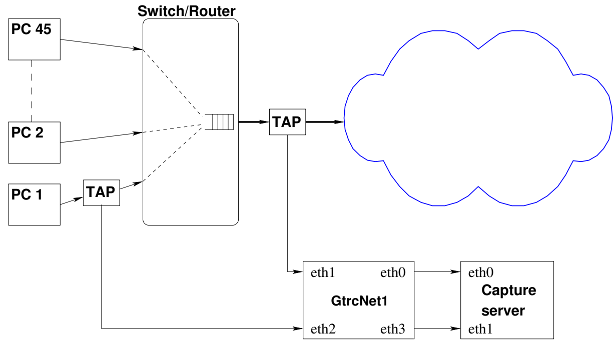

Figure 3: Experimental topology

particular source among the N, which can be viewed as a probing source. The output of the buffer constitutes a 1 Gbps bottleneck whose traffic is integrally captured by entry 2 of the *Metroflux* system. After identification of the packets pertaining to the probing flows, the time-stamp difference between entries 1 and 2 correspond to the queuing delay introduced by the instantaneous buffer state.

Traffic model, metrics and measurements. Considering each source as a ON/OFF model, we make vary :

- the statistics of the ON (OFF) durations,
- the number of competing sources,
- the mean source throughput,
- the sources' correlation.

For each configuration, we capture the delay time series corresponding to the necessary time for probing packets to go through the buffer. The probing source can be used to trigger the buffer occupancy measurement according to a periodic clock (uniform sampling), or to adapt the measurement rate to the instantaneous traffic load (adaptive sampling).

Expected analyses from the captured traces. We can think of several approaches to characterize the load dynamics of the bottleneck queue. One of them is motivated by the works achieved independently by I. Norros and M. Mandjes. The raised question concerns the buffer load distribution in the case of fBm-like input traffic. Whereas the first study predicts a Weibul distribution, the second announces a heavy tailed distribution. Both theoretically sound, only the experimental conditions should be determinant, and we would like to verify which conjecture holds in practice and to which extent.
A recent theoretical work by M. Roughan defines stringent Heisenberg uncertainty principles that should limit the expectable precision of some experimental measures. That is notably the case for the packet processing delay, transmit time, propagation delay and queuing delay. Our experimental plate-form can provide with a flexible and accurate instrumentation to assess the practical validity of such theoretical predictions.