Do you know what you are generating?*

Alberto Dainotti and Alessio Botta†
Dipartimento di Informatica e Sistemistica,
University of Napoli, Italy.
{alberto,a.botta}@unina.it

Antonio Pescapè
Dipartimento di Informatica e Sistemistica,
University of Napoli, Italy.
pescape@unina.it

ABSTRACT

Software-based traffic generators are commonly used in experimental research on computer networks. However, there are no much studies focusing on how such instruments are accurate. Here we start a discussion reviewing the problem of using software-based traffic generators over common hardware/software, highlighting interesting issues that pose some threats to common beliefs. We started comparing the operator-requested traffic profile against the real behavior of commonly used software-based traffic generators. We aim at performing tests under different conditions and looking both at packet/bit rate and inter-packet time distribution. Preliminary results show notable differences in some cases, opening the way to interesting discussions and further investigations.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Miscellaneous

Keywords

Measurement, Traffic Generation

1. INTRODUCTION

In the study and research on computer networks the problem of generating synthetic traffic (i.e. not generated by common applications or by the control plane) for several tasks has always existed. Such generation is usually done by using active measurement tools, with the purpose to perform a measurement experiment. The recent literature witnesses the necessity to perform experiments with tools that are able to replicate several statistical properties of real traffic [1] [2] [3]. Traffic generators are implemented as either hardware or software platforms. Typically, the former are more precise and expensive, reach higher performance, and are generally developed by the Industry[1]; the latter are less precise, cheaper, generally developed by Research units or Universities[2], and often released as open-source. Most of the times, researchers use software platforms. This is mainly due to their flexibility (e.g. easy deployability of several nodes, possibility to modify the code for a specific research purpose,...). Moreover, they allow to make more realistic experiments by testing actual implementations running on top of real OSes and network protocol stacks. On the other hand, compared to hardware traffic generators (seen as measurement instruments), software platforms can not provide detailed datasheets containing certified information on the tool capabilities because their metrological properties (e.g. accuracy) depend on the COTS (Commercial Off-The-Shelf) hardware used, the OS adopted, and the status of the traffic generation machine [4]. Therefore, without a preliminary analysis of the metrological properties of the used tool, the reference (i.e. injected input) remains uncertain and consequently results could be invalidated.

2. MOTIVATION

How much does networking research rely on traffic generators? A (cleaned) query for the string "traffic generator" on the IEEE search engine reported a number of matching published papers that, from an average of 30 per year until 2001, jumps to about 150 since 2002, with a peak of about 200 in 2005. We made a simple search on both Google and Google Scholar (as of July 2007) of works using some of the widely adopted software traffic generators we test in our research. We found 245 works (papers, tutorial, technical reports ...) taking into account MGEN[3], 150 citing RUDE/CRUDE[4]. D-ITG[5], our traffic generator, reports about 150 hits

*This work has been partially supported by CONTENT EU NoE, OneLab, and NETQOS EU projects.

†PhD Students.

CoNEXT'07, December 10-13, 2007, New York, NY, U.S.A.
CONEXT '07 New York, USA - 10-13 December 2006

[1] Agilent HP, Candelatech, IXIA, Skaion, Omnicor, Spirinet
[2] Mgen, Rude/Crude, TG, D-ITG,...
[3] http://cs.itd.nrl.navy.mil/work/mgen/index.php
[4] http://rude.sourceforge.net/
[5] http://www.grid.unina.it/software/ITG

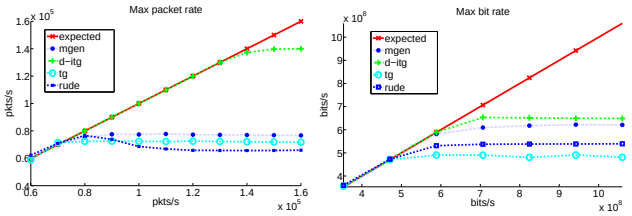

Figure 1: Generated packet- and bit-rate.

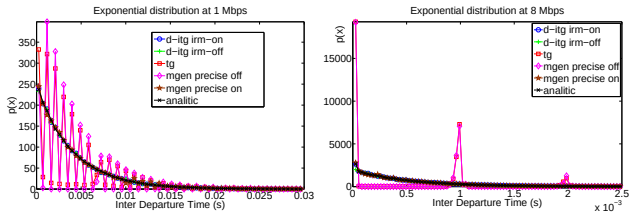

Figure 2: PDF of requested and generated IDT.

on Google and Google Scholar. Generally, traffic generation platforms take in consideration models of real traffic, but there is often no knowledge of the accuracy in the generation of synthetic traffic: this heavily affects results of a research experiment. The following are only few examples: (i) available bandwidth measurements using synthetic UDP and TCP cross-traffic; (ii) testing the benefits of using Diffserv and IntServ architectures; (iii) active measurements of links delay using Poisson traffic; (iv) testing statistical based anomaly and intrusion detection systems; (v) comparing different transport protocols with realistic loads and background traffic; (vi) analyzing queue performance with realistic traffic. However, in the research field of active measurements on networks, the attention is usually paid only to the output, intending it as the results provided by the active tool, whereas few or no considerations at all are made on the quality of the input. Considering the plethora of results from the above mentioned networking fields, the idea at the base of this work comes out from the following considerations/questions: 1) Do you know what you are generating? 2) Do you trust your results (on the base of what you generated)?

3. GENERATING NETWORK TRAFFIC

We are testing a number of traffic generators over a simple and real network testbed. We are analyzing their ability to actually generate the imposed bitrate and packet rate, and to accurately reproduce marginal distributions for packet IDT that are commonly used in scientific experiments (e.g. exponential inter-departure times). In the following, we will give preliminary results of such experimental evaluation. Moreover, we foresee to evaluate the impact of noticeable divergences between the traffic profiles expected and those really produced also to evaluate if such neglected differences had an impact on past literature or future research.

Accuracy of average bitrate and packet rate: In Fig. 1 (left) we reported the average packet rate that the considered generators were able to inject into the network. As shown, D-ITG starts to deviate from the ideal behavior at about 130Kpps, while the best performing among the remaining tools (i.e. MGEN), is not able to generate more than 80Kpps. This witnesses that D-ITG can be used to generate a higher packet rate than all the other tools. Concerning the bitrate, Fig. 1 (right) shows a more balanced situations. Indeed, D-ITG is still able to achieve an higher bitrate, i.e. 630Mbps, but the difference with the other tools is not as high as in the case of the packet rate.

Accuracy of the imposed IDTs: In Fig. 2 we reported the PDFs of the IDTs obtained when generating 1 Mbps(left) and 8 Mbps(right) of UDP traffic, with Exponential IDT and constant PS (equal to 512 Bytes): TG generates packets only at every integer multiple of $1ms$ being not able to reproduce an accurate PDFs; the same applies also to MGEN when the *precise* feature is disabled (i.e. if such an analysis is not performed, the researcher could believe to be testing the network with Poisson traffic while this is not the case).

4. DISCUSSION AND CONCLUSION

This research activity aims at pointing the attention of the research community to what we consider an underestimated subject. Can we consider the synthetic traffic generators which we currently use in several areas of experimental network research as real measurement instruments? If so, what is their real performance under different conditions, and how accurately do they replicate the user-requested statistical models? Moreover, how heavily may these neglected differences, between real operational and expected behaviors, affect the results of experimental activity? Did past scientific literature produce wrong results by relying on them? What is the state of the art of software-based traffic generation? And what can we do to improve it? In this abstract we have heavily summarized the background and motivation of our ongoing research on this subject and we reported some of the first results. We hope that proposing these themes in a poster session will allow us to both stimulate productive discussion and collect precious feedback for our research work.

5. REFERENCES

[1] F. Dressler, "Policy-based traffic generation for IP-based networks", Poster at *INFOCOM 2006*

[2] K. V. Vishwanath, A. Vahdat, "Realistic and responsive network traffic generation" *ACM SIGCOMM Computer Communication Review*, Volume 36 , Issue 4 (October 2006) pp. 111-122

[3] A. Botta, A. Dainotti, A. Pescapé, "Multi-protocol and multi-platform traffic generation and measurement", *INFOCOM 2007 DEMO Session*

[4] M. Paredes-Farrera, M. Fleury, M. Ghanbari, "Precision and Accuracy of Network Traffic Generators for Packet-by-Packet Traffic Analysis", *2nd International IEEE TridentCom Conference*, pp. 32-27, Barcelona, Mar. 2006.