On the Impact of the Flow-Size Distribution's Tail Index on Network Performance with TCP Connections

Oana Goga
CNRS and UPMC Sorbonne Universités, France
oana.goga@lip6.fr

Patrick Loiseau
University of California, Santa Cruz, CA, USA
ploiseau@soe.ucsc.edu

Paulo Gonçalves
INRIA and Université de Lyon / ENS Lyon, France
paulo.goncalves@ens-lyon.fr

ABSTRACT

In this paper, we study the impact of the flow-size distribution on network performance in the case of a single bottleneck with finite buffer. To tackle the case where flows are transmitted with the TCP protocol, we use real experiments and ns-2 simulations. Our preliminary results show that the distribution's tail index impacts the performance in a more complex way than what is reported in existing literature. In particular, we exhibit situations where a heavier tail gives better performance for certain metrics. We argue that a main cause of our observed results is the transient behavior at the beginning of each flow.

Keywords

Heavy-tailed distribution, network performance, TCP

1. INTRODUCTION

1.1 Motivations and related work

It is well established that the distribution of flow sizes in the Internet is heavy-tailed [3]. Following up this important discovery, in the last 15 years or so, significant research efforts have been devoted to understanding the impact of this property on network performance, mainly in the case of a single bottleneck. Experimental and theoretical works using an open-loop approach have concluded that heavy tails degrade performance for large buffers, whereas performance is insensitive to the tail for small buffers [4, 6, 8]. However, most of today's Internet traffic uses the TCP protocol, which is based on a closed-loop mechanism. In this case, the simulation-based study of [9] showed that, with their particular setting, heavier tails (*i.e.,* smaller tail indices) degrade the performance for all metrics (loss, delay, throughput). Most of the rest of the literature concluded, based on various models, that first-order performance metrics (*e.g.,* mean flow-rate) are insensitive to the flow-size distribution [1, 2]. However, these models use simplifying assumptions, notably on bandwidth sharing, that do not account for the transient behavior at the beginning of each TCP flow. Finally, numerical methods based on more complex models have been proposed to evaluate performance metrics given the input conditions [5]; but they do not provide qualitative insight to understand the impact of the flow-size distribution.

1.2 Contributions

In this paper, we study the impact of the flow-size distribution's tail index in the case of a single bottleneck of finite buffer, with the TCP protocol. To grasp the full complexity of the problem, we use experiments on a real (but fully controlled) network testbed [7] rather than models based on simplifying assumptions. We also use ns-2 simulations to complement our study. We report preliminary results that contrast with existing literature by showing that:

- the impact of the tail index on performance depends on many parameters and not only on the buffer size; and
- there exists situations where a heavier tail gives better performance for certain metrics.

To the best of our knowledge, this last effect has not been reported in previous literature. We argue that an important element to interpret these results is the transient behavior, *i.e.,* the flow's behavior during slow-start, which is not taken into account in the aforementioned models.

2. SETTING

We use a butterfly topology with the same number N_{src} of sources and destinations. Experiments on a real network and ns-2 simulations use the same setting with $N_{\text{src}} = 45$ and 500 sources respectively. Each source behaves as an ON/OFF source: it alternates between flow transmission using the TCP protocol (with Reno variant) and idle (OFF) period. To avoid multiple congestion points, each source always sends to the same destination and there is no bottleneck on the return path (for the ACKs transmission). The minimal RTT (excluding queueing delay) is 10 ms for each pair source-destination. The bottleneck queue uses a Drop Tail policy with finite buffer of size B which is 96 or 896 pkts in the experiments and 100 or $1,000$ pkts in the simulations. These values correspond to maximal queueing delays of 1.15, 10.75, 1.2 and 12 ms, respectively. The capacity of both the source links and the bottleneck link is 1 Gb/s. Each experiment lasts 2 hours and each simulation 1 hour.

We use a Pareto distribution for the flow-size distribution, of mean $1,000$ pkts and of tail index α varying across experiments to assess its impact. In all the experiments/simulations, we use an exponential OFF-time distribution of mean 0.2 s. Note that due to the closed-loop mechanism of TCP, the load may vary between the experiments [11].

An important parameter in our study is the TCP tuning parameter which controls whether the `ssthresh` parameter is cached from one TCP connection to the next connection with the same source and destination. The `ssthresh` parameter determines the maximal congestion window of the slow-start. When it is reached, the connection goes to the congestion avoidance phase, whereas if a loss occurs before

it is reached, then the congestion window at this loss event becomes the new **ssthresh**. If it is cached, then it can only decrease from one flow to the next flow. In our experiments, there is a non-negligible loss rate. Then, the **ssthresh** rapidly stabilizes around 10 pkts for each source-destination pair, which significantly limits the slow-start phase. We refer to this situation where the **ssthresh** is cached as "without slow-start". It is the default situation in our linux implementation. We also perform experiments with the caching manually disabled (*i.e.,* "with slow-start"). Both situations may represent different real-life cases. Ns-2 simulations are performed with disabled caching.

For the experiments on a real network, the performance results are derived from the analysis of synchronized packet-level captures at the input and at the output of the limiting buffer. For the ns-2 simulations, we use monitoring directly in the code. Error bars displayed correspond to the standard deviation on the estimation of the mean, estimated via bootstrap techniques on each experiment/simulation. Many are so small that they are not visible on the plots. Note however that for α close to one, the results across different experiments may have variability not included in these error bars, due to the natural difficulty to impose the mean of the flow-size distribution.

3. RESULTS

Fig. 1 presents the results obtained from the experiments on a real network with $N_{\text{src}} = 45$ sources. The top plots show the sensitivity of buffer-level performance metrics (loss rate, mean delay and throughput) to the tail index α. We observe that the impact of α not only depends on the buffer size (as shown in previous literature), but also on the presence/absence of the slow-start phase. Moreover, the impact varies from metric to metric.

More precisely, in the case "without slow-start", the performance in terms of loss rate and mean delay degrades with heavier tails, whereas the mean throughput increases with heavier tails which corresponds to a performance improvement. In the case "with slow-start", most results are inverted, except for the throughput which appears almost insensitive to the tail index for a small buffer ($B = 96$ pkts). In particular, in terms of loss rate and mean delay, a heavier tail turns out to give better performance. This non-intuitive result has not been reported in previous literature.

To interpret the results, the bottom plots of Fig. 1 show the mean rate achieved by a flow, as a function of its size (in loglog scale). In both cases, it is a strongly non-flat function, which stabilizes only for very large flow sizes. This is due to the effect of the transient behavior of each flow, which has an important impact on the mean rate even for large flows (see also [10]). This transient behavior is longer and smoother "without slow-start"; it is faster and sharper "with slow-start" which explains the form of the bottom plots of Fig. 1. In the case "with slow-start", due to lower loss rate and mean delay, large flows also have a higher mean rate for smaller α's. Then, the performance results are mainly driven by an intricate combination of the shape of the mean flow-rate curve and of the flow-size distribution; which imposes the load. This load impacts the loss rate and delay, which in turn impacts the mean flow-rate curve, in a closed-loop manner. It appears from this interpretation that the transient behavior of each flow is largely responsible for the impact of the flow-size distribution observed on performance.

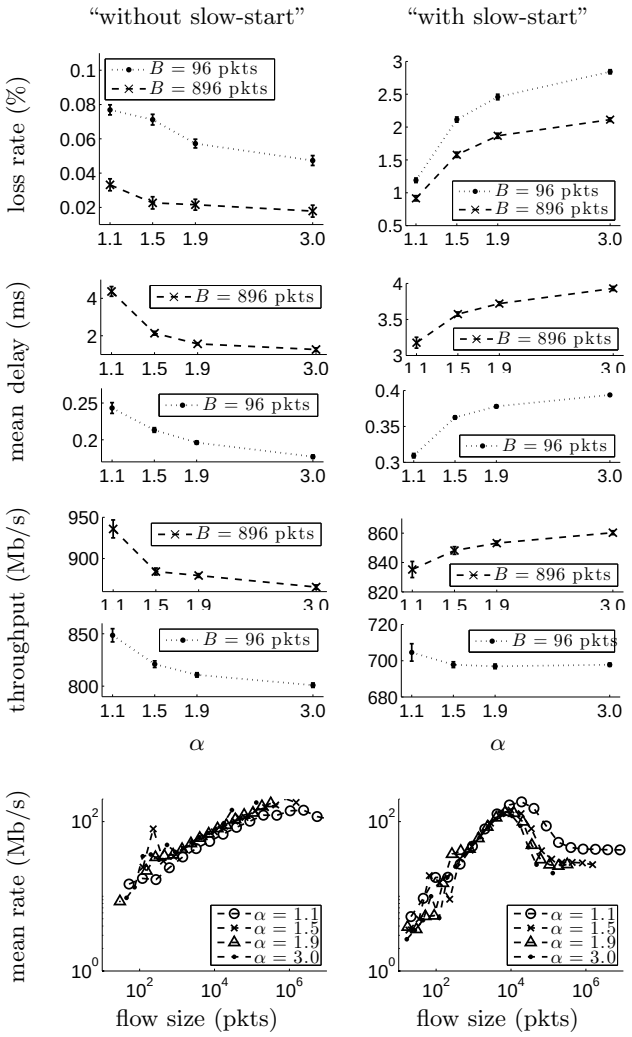

Figure 1: Performance metrics results as a function of α, from the experiments with $N_{\text{src}} = 45$ sources. Whenever necessary, the curves corresponding to the different B's have been split and the scales adapted to improve readability. Flow rates (bottom plot) are presented for $B = 896$ pkts.

To corroborate this interpretation, Fig. 2 presents the results obtained with simulations using $N_{\text{src}} = 500$ sources, in the case "with slow-start". In the case where the buffer is small ($B = 100$ pkts), the transient regime is very limited. Therefore the mean flow rate appears almost flat (bottom-left plot), and the buffer-level performance metrics are roughly insensitive to the tail index of the flow-size distribution (except for $\alpha = 1.1$, due to the difficulty to impose the mean flow size for such a small α). However, with a larger buffer ($B = 1,000$ pkts), the effect of the longer transient regimes becomes important (bottom-right plot, compare also to the right-side of Fig. 1 corresponding to the same case with less sources), and performance becomes sensitive to α. Interestingly though, the evolution is different from the case of Fig. 1-(right) with $N_{\text{src}} = 45$. As mentioned previously, we believe that it is the result of a complex combination between the mean rate as a function of the flow size, and the flow-size distribution. In particular, it appears that the

most "aggressive" flows in terms of average rate are not the largest ones, but the ones of "medium size". This suggests that a key to fully understand the impact of the flow-size distribution in complex situations involving TCP connections with transient regimes is to look at the impact of the distribution as a whole, rather than the impact of its tail index only. We leave such a complete study as future work.

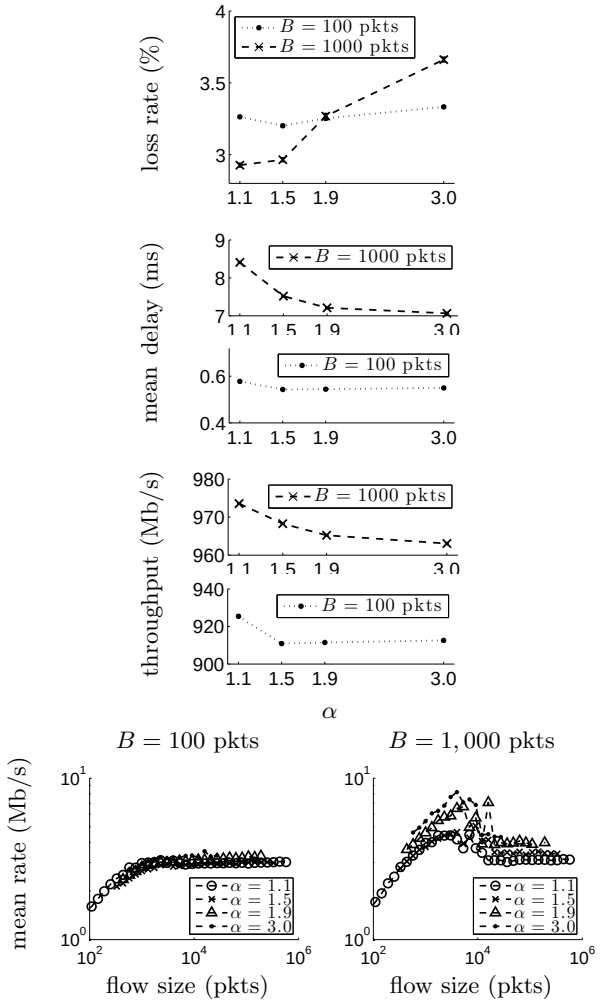

Figure 2: Performance metrics results as a function of α, from the simulations with $N_{\mathbf{src}} = 500$ sources. The simulations are performed in the situation "with slow-start". Whenever necessary, the curves corresponding to the different B's have been split and the scales adapted to improve readability.

4. CONCLUSION

In this work, we presented experimental and simulation results on the impact of the flow-size distribution's tail index on the performance of a single bottleneck with TCP connections. Our preliminary results show that the impact of heavy tails on performance is more complex than what described in previous literature: it can be negative, negligible, or even positive depending on many parameters such as buffer size, metric considered (loss rate, mean delay or throughput) and TCP tuning parameters (`ssthresh`).

We argue that a main element to understand these results lies in the transient behavior of each flow, which relates to the slow-start mechanism. Therefore, our work identifies the non-stationarity of a flow's transmission as an important feature (which is absent from all tractable models proposed up to now) to understand the performance of TCP connections.

Though our study shows a clear impact of the flow-size distribution's tail index on performance, a more complete study is required to confirm our interpretations and to find meaningful parameters to intuitively understand better the impact of this distribution. We believe that a key toward this goal is to consider the impact of the whole distribution rather than its tail index only.

5. ACKNOWLEDGMENTS

We are grateful to the ALADDIN-G5K initiative whose engineers provided us with full support to perform experiments on the Grid'5000 infrastructure. We thank C. Dovrolis and R. Prasad for sharing some of their ns-2 code with us. Part of this work was done while the first two authors were with INRIA and Université de Lyon / ENS Lyon, France.

6. REFERENCES

[1] A. Arvidsson and P. Karlsson. On traffic models for TCP/IP. In *Proc. of ITC-16*, 1999.

[2] S. Ben Fredj, T. Bonald, A. Proutiere, G. Régnié, and J. W. Roberts. Statistical bandwidth sharing: a study of congestion at flow level. In *Proc. of ACM SIGCOMM '01*, pages 111–122, 2001.

[3] M. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Trans. Netw.*, 5(6):835–846, Dec. 1997.

[4] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Trans. on Netw.*, 4(2):209–223, 1996.

[5] M. Garetto and D. Towsley. An efficient technique to analyze the impact of bursty TCP traffic in wide-area networks. *Perf. Eval.*, 65(2):181–202, 2008.

[6] M. Grossglauser and J.-C. Bolot. On the relevance of long-range dependence in network traffic. *IEEE/ACM Trans. Netw.*, 7(5):629–640, 1999.

[7] P. Loiseau, P. Gonçalves, R. Guillier, M. Imbert, Y. Kodama, and P. Vicat-Blanc Primet. Metroflux: A high performance system for analyzing flow at very fine-grain. In *Proc. of TridentCom*, pages 1–9, 2009.

[8] M. Mandjes and N. Boots. The shape of the loss curve and the impact of long-range dependence on network performance. *AEU - Int. J. Elec. Commun.*, 58(2):101–117, 2004.

[9] K. Park, G. Kim, and M. Crovella. On the effect of traffic self-similarity on network performances. In *Proc. of SPIE Int. Conf. on Perf. and Control of Netw. Syst.*, pages 296–310, Nov. 1997.

[10] R. S. Prasad, C. Dovrolis, and M. Thottan. Router buffer sizing revisited: the role of the output/input capacity ratio. In *Proc. of ACM CoNEXT '07*, pages 15:1–15:12, 2007.

[11] B. Schroeder, A. Wierman, and M. Harchol-Balter. Open versus closed: A cautionary tale. In *Proc. of NSDI '06*, pages 239–252, 2006.