Cross-Traffic: Noise or Data?

C. Blake, D. Katabi, S. Katti
Computer Science and Artificial Intelligence Laboratory @ MIT

I. Overview

Many path measurement techniques have tried to eliminate the effects of cross-traffic or take cross-traffic into account via fluid models [4], [6], [7], [8]. In contrast, this work unravels the structure of cross-traffic and shows a way to leverage it. Specifically, we show how to use cross-traffic structure to discover the capacity and relative location of multiple upstream congested links traversed by TCP flows. We also describe a tool, `multiQ`, for extracting this information from only a receiver side `tcpdump`.

II. Cross-Traffic in the Internet

This work focuses on TCP flows as candidates for passive probes into the state of the network. Yet, stalled or very short TCP connections are poor probes. To filter out such inappropriate flows, we only consider **significant flows**, which we loosely define to be TCP connections that last for at least a couple of seconds, achieve an overall average packet rate of at least 20 pps ($\approx$ 2 pkt/RTT), and contain at least one 1500 byte packet.

We define a **cross-traffic burst** to be the traffic that intervenes between two consecutive packets of a significant flow. At each hop the individual burst between a specific packet pair may differ. However, we seek to understand the probabilities of various amounts of traffic intervening between a pair of packets in a significant flow at a congested router. Thus, we are only interested in characterizing the probability distribution of intervening burst sizes. Cross-traffic bursts are interesting for passive measurement techniques because queueing translates them into inter-arrival times observable at downstream receivers.

To assess the distribution of cross-traffic in the Internet, we studied 300 million packets in 162 NLANR traces [1] that spanned several months of traffic on one OC12 and eleven OC3 links. We identified 28,000 significant flows. For each pair of packets in a significant flow, we computed the intervening cross-traffic burst at the link where the trace is taken. This creates one sample burst size. The distribution of all bursts is shown in Figure 1a. Note the uniform gaps of 1500 bytes between the sharp modes in the burst distribution. This distribution is intriguing because of its surprising regularity. It contains sharp modes separated by intervals of 1500 bytes. To explain this regularity we examined the packet size distribution for the traces illustrated in Figure 1b. (Note the similarity with the packet size distribution reported in [9].) The packet size distribution shows

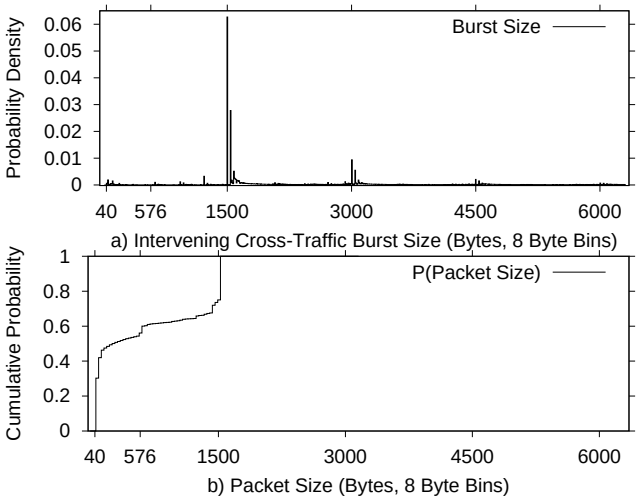

Fig. 1. a) Cross-traffi c between consecutive packets in a signifi cant fbw has a distribution with sharp mode structure. b) The CDF of packet size reveals frequencies of 40 and 1500 byte packets.

dominant frequencies of 40 and 1500 byte packets. There are many other sizes but none of them is highly pronounced. Hence, it is expected that the structure in the burst distribution will stem from 40 byte and 1500 byte packets. Yet, because of the very large difference in size between 1500 and 40 byte packets, the small packets only broaden the modes in the burst size distribution. While we cannot ensure that the cross-traffic burst distribution seen in these traces is representative of cross-traffic everywhere in the Internet, we believe the diversity and large size of our data makes such a generalization plausible.

Since the large scale structure in cross-traffic bursts has 1500 byte gaps, one would expect packet inter-arrival distributions to have modes with gaps equal to the transmission time of a 1500 byte packet on congested links.

III. PDF of Inter-Arrival Times in a TCP Flow

We motivate our work by discribing the outcome of a simple experiment. We examine the path connecting two machines. The first machine is at CMU and has an access link of 10 Mb/s, whereas the second is at CCICOM and has a 100 Mb/s access link. We first download a long file from CCICOM to CMU, and look at the packet inter-arrivals at CMU using `tcpdump`. We plot the distribution of these interarrivals in Figure 2a. The distribution shows a single spike at 1.2 ms, which is the transmission time of a 1500 byte packet on a 10 Mb/s link. Next, we repeat the experiment along the *reverse path*; we download a long file

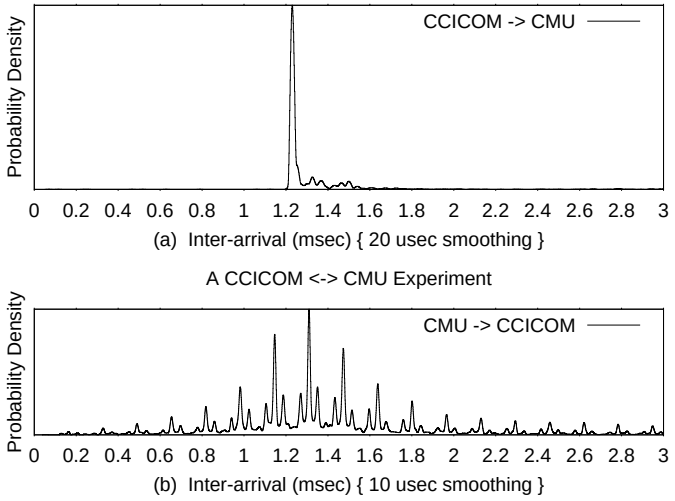

Fig. 2. Inter-arrival PDFs for CCICOM-CMU path in both directions.

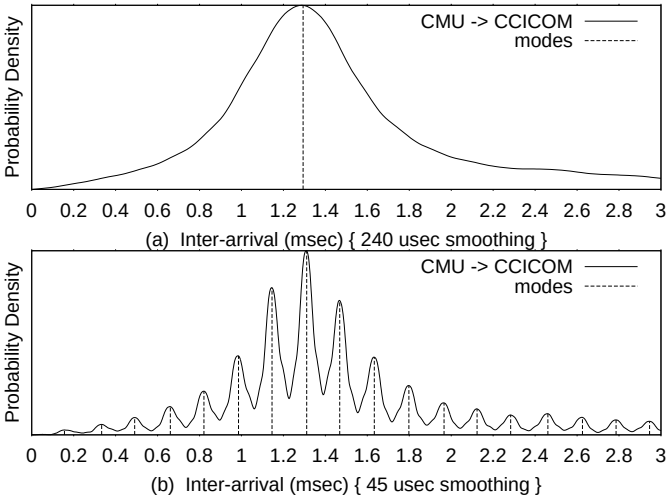

Fig. 3. The data from Figure 2b at two smoothing scales.

from CMU to CCICOM and plot the packet inter-arrival distribution as seen by `tcpdump` at CCICOM. The resulting distribution, illustrated in Figure 2b, has an interesting structure. The envelope of the distribution is centered near 1.2 ms, which is the time to send a 1500 byte packet on the upstream 10 Mb/s link. This envelope is modulated with sharp spikes separated by a gap of 0.12 ms, which is the transmission time of a 1500 byte packet on a 100 Mb/s link. Said differently, the inter-arrival PDF reveals that packets faced queuing on both access links. Further, the envelope of the PDF describes the first congested link along the path, whose output gets modulated by the next congested link.

This simple experiment shows that passive observations of packet inter-arrivals at the receiver convey information about the capacity and relative order of congested links along the path traversed by the flow. The example PDFs in Figures 2a and 2b are not the only possible cases. However, inspection of inter-arrival PDFs for over 200 different paths shows that it is typical of all these PDFs to exhibit modes separated by the transmission time of a 1500 byte packet on a well-know link capacity. For lack of space we do not show these PDFs.

IV. MULTIQ

We can automate the analysis of the last section. What `multiQ` does is estimate the inter-arrival density at a progression of smoothing scales corresponding to the known set of common link speeds. At each smoothing scale `multiQ` constructs kernel density estimate of the PDF, and scans it for statistically significant modes as in Figure 3 (which plots the same data in 2b at a different smoothing scale). The *gaps* between these modes are then computed. The distribution of these mode gaps also has modes of its own. The location of the first mode in the mode gap distribution is in fact the time for a 1500 byte transmission on the router singled out by the smoothing scale.

V. CONTRIBUTIONS

Our work has the following contributions.

1. We uncover the distribution of cross-traffic bursts which intervene between consecutive packets in a significant TCP connection, and discuss the implications of this distribution for passive measurements.

2. We present a method for inferring the capacity of the congested links along a path and their relative order. This differs from prior work on capacity estimation [2], [7], [5] which discovers the capacity of one or more links along a path, without returning information about their state of congestion. In comparison with Tulip [8] and BFind [3] which discover the congested links along a path, our approach is completely passive; it does not generate any extra load nor does it require ICMP support at routers.

3. We built a tool, multiQ, which automatically gleans information about the capacity and order of upstream congested links from the `tcpdump` at the receiving side.

REFERENCES

[1] National laboratory for applied network research (nlanr). http://www.nlanr.net/.
[2] pathchar. ftp://ee.lbl.gov/pathchar.tar.Z.
[3] A. Akella, S. Seshan, and A. Shaikh. An empirical evaluation of wide-area internet bottlenecks. In *Proc. IMC*, 2003.
[4] M. Allman and V. Paxson. On estimating end-to-end network path properties. In *The Proc. of ACM SIGCOMM '99*, Aug. 1999.
[5] R. Carter and M. Crovella. Measuring bottleneck link speed in packet-switched network. Technical Report TR-96-006, Boston University, Mar. 1996.
[6] A. B. Downey. Using pathchar to estimate internet link characteristics. In *The Proc. of ACM SIGCOMM '99*, Aug. 1999.
[7] K. Lai and M. Baker. Nettimer: A tool for measuring bottleneck link bandwidth. In *Proc. of USENIX*, Mar. 2001.
[8] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson. User level internet path diagnosis. In *Proc. ACM SOSP*, Oct. 2003.
[9] C. Shannon, D. Moore, and K. Claffy. Beyond folklore: Observations on fragmented traffi c. In *IEEE/ACM Transactions on Networking*, Dec. 2002.