Measuring Interdomain Congestion and its Impact on QoE

Amogh Dhamdhere
CAIDA / UC San Diego
amogh@caida.org

Matthew Luckie
CAIDA / UC San Diego
mjl@caida.org

David Clark
MIT
ddc@csail.mit.edu

kc claffy
CAIDA / UC San Diego
kc@caida.org

ABSTRACT

We have developed a method to localize and quantify interdomain congestion in the Internet. Our Time Sequence Latency Probes (TSLP) method depends on two facts: Internet traffic patterns are typically diurnal, and queues increase packet delay through a router during periods of adjacent link congestion. Repeated round trip delay measurements from a single test point to the two edges of a congested link will show sustained increased latency to the far (but not to the near) side of the link. We are designing and implementing a system for network-wide measurement of congestion using the TSLP method. We plan to support QoE measurements on this system to complement our measurement of QoS metrics such as delay and loss rate.

1. INTRODUCTION

Unlike traffic congestion on links within a single network (AS), where responsibility for resolving the congestion unambiguously belongs to that network, congestion on AS interconnection links (or *interdomain congestion*) may reflect a peering dispute, accompanied by finger-pointing over which network should pay to upgrade the link to handle the traffic demand. The two primary forms of interconnection are *transit*, when one AS sells another ISP access to the global Internet, and *peering*, when two ISPs interconnect to exchange customer traffic. The historical basis for settlement-free peering was a presumed balance of value to both parties. Peering disputes arise when one party believes the exchange is no longer beneficial to them. Historically, peering disputes were between large transit networks (e.g. [7, 11]) where one party would fall out of compliance with the agreement and be disconnected by the other party until a new agreement was reached. More recent peering disputes are fueled by exploding demand for high-bandwidth content (e.g., streaming video), and growing concentration of content among a few content distribution networks (e.g. [1–4, 8, 9, 13]), some large and sophisticated enough to adjust loading (and thus congestion levels) on interconnection links [6, 10]. Many disputes do not lead to disconnection but stalled negotiation about who should pay for installation of new capacity to handle the demand, leaving the congested link as an externality for all users of the link until the dispute is resolved.

Congestion on interdomain links, especially between large transit providers, can affect a large number of end-to-end paths that traverse those links. A congested link causes the buffer at the link to fill up, introducing additional queueing delay that increases the round-trip time of all connections traversing that link. Additionally, a persistently congested link introduces a non-negligible loss rate. Finally, a link that is running at or close to capacity has, by definition, low (or zero) available bandwidth. This means that connections traversing such a link must obtain bandwidth by pushing other connections out of the way. Delay, loss, and throughput all contribute to the Quality of Experience of end-users depending on the specific applications they use. For example, streaming video will perform poorly on a lossy path with low available bandwidth, causing the video to buffer or switch to a lower quality encoding. We believe that inferring the presence of congested links and how they impact QoS parameters such as latency, loss and available bandwidth is important for understanding application QoE.

2. TIME SERIES LATENCY PROBES

The idea behind the time-sequence latency probes (TSLP) method is to frequently repeat round trip time (RTT) measurements from a vantage point (VP) to the *near* and *far* routers of an interdomain link. The measured RTTs are a function of the queue lengths of the routers on the forward and reverse paths: as queue lengths increase, so does RTT. When RTTs increase to the far router but not to the near router, we infer that a queue between these two routers induced the delay.

If a link is so busy that a tail-drop queue is always close to full, a time series of RTT measurements to the far router will approximate a square wave, with the minimum RTT during the low state reflecting probes that did not experience delay, and the minimum RTT during the high state reflecting probes consistently encountering a queue close to full. Queue lengths are finite, limiting the delay contributed by any one queue, reflected by the top of the square wave. Figure 1 shows such an RTT pattern on a peering link between Comcast and Cogent; the minimum RTT measured every five minutes to the Cogent router increased from 20ms to 70ms for 14-18 hours per day. We also probed every second to observe packet loss across this link, which we only observed in periods where we also observed increased RTT. We hypothesize that the increasing loss rate correlates with increasing demand on the link, and that the width of the period with elevated delays reflects the length of time the link was congested. The height of the elevated period is not an indication of the *degree* of congestion, but rather the size of a queue in a router serving the interdomain link.

3. RECENT EPISODES OF INTERDOMAIN CONGESTION

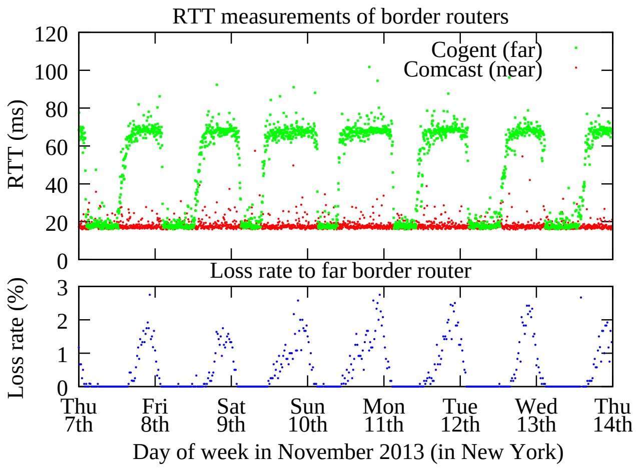

Figure 1: Congestion on an interdomain link between Comcast and Cogent, measured from a VP within Comcast. The RTT to the Cogent (far) router increases from 20ms to 70ms while the RTT to the Comcast (near) router is stable at 20ms. The approximate square wave indicates the queue is always close to full when the RTT increases to 70ms. The loss rate from the Cogent router increases after this level shift occurs, as the load on the link continues to increase.

We describe next two examples of interdomain congestion which we have measured using our congestion measurement system. The first case, shown in Figure 2, shows RTTs to the near and far end of an interdomain link between network A and B, measured from a VP in network A during one week in April 2015. We see evidence of congestion for 4-5 hours every day during April and May 2015 on this link.

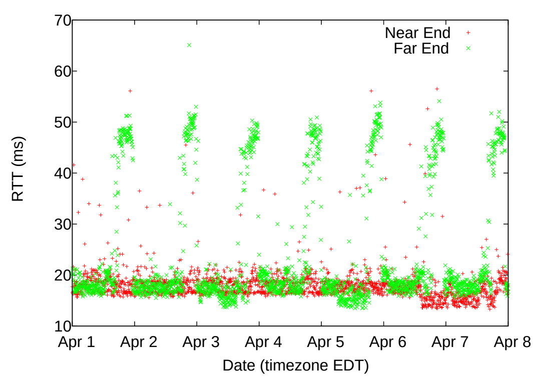

Figure 2: Congestion on an interdomain link between network A and B, measured from a VP within network A, in April 2015. TSP probes show evidence of congestion for approximately 4-5 hours during peak times in the local time zone of the VP.

Figure 3 shows RTTs to the near and far end of an interdomain link between network C and network D, measured from a VP in network C during one week in May 2015. We observe that on 4 days during the last week of May 2015, the RTT to the far end of the link increased for a short duration (less than 2 hrs) during peak hours. This example is quite different from the network A-B case, as the congestion appears for only 4 days in May 2015, and lasts for a short period of time.

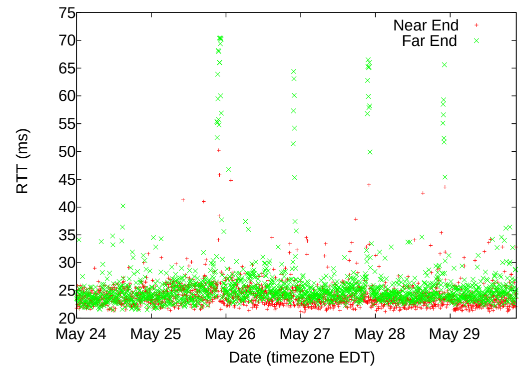

Figure 3: An interdomain link between network C and D, measured from a VP within network C, in May 2015. TSP probes show moderate congestion on 4 days in 2015 for 1-2 hours during peak times in the local time zone of the VP.

An open question is how these different types of congestion (sustained for several hours and recurring every day vs. short-lived, one-off episodes) affect the Quality of Experience for users running different applications on paths traversing these congested links. Do transient congestion episodes lead to a perceptible change in user QoE? The measurement system we are developing will enable a deeper investigation of interdomain congestion episodes and potential impacts on user QoE.

4. SYSTEM ARCHITECTURE

We are in the process of designing and implementing a comprehensive congestion measurement system. Figure 4 shows an overview of the various components of the system. The Vantage Points are measurement hosts that are responsible for performing the actual measurements. Currently our system uses VPs from CAIDA's Ark [5] infrastructure, which we are expanding to include Bismark [12] VPs. The VPs run several processes concurrently, among which is a topology discovery process aimed at discovering the interdomain links of the network hosting the VP and visible from that VP. Our backed system processes the data collected by this topology discovery process to determine the set of interdomain links to probe from each VP. This *target set* is provided as input to the TSP process that performs TTL-limited probing from the VPs. The data management module in the backend system collects TSP data from the VPs, indexes the data into databases, generates longitudinal time series, and applies various time-series analysis techniques (level-shift detection and frequency-domain analysis) to infer congestion at interdomain links.

We are developing an *alert system* to generate alerts in close to real time when our analysis shows evidence of congested interdomain links. The alert system will trigger additional *reactive measurements* from the VPs. The reactive measurement system is driven by a centralized measurement scheduler that dispatches measurements to VPs based on

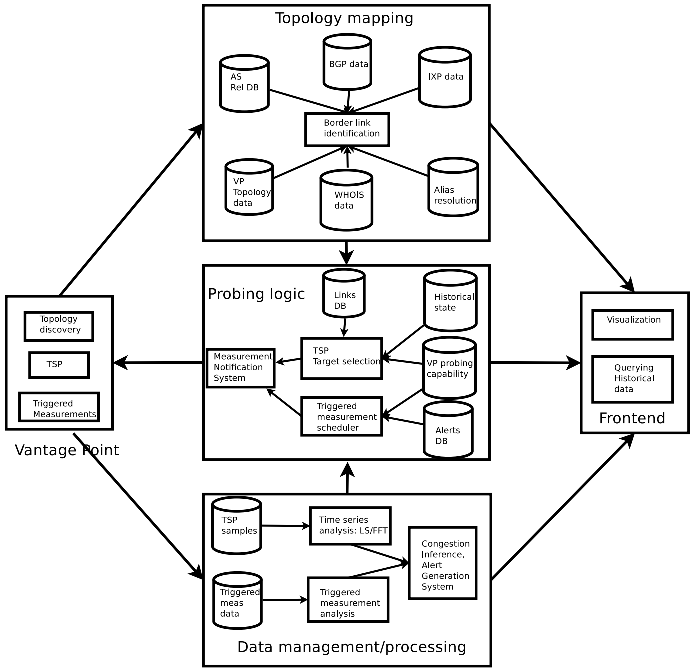

Figure 4: Overview of our congestion measurement system.

generated alerts and external criteria (such as the probing capability and probing budget of the VP). A client running on the VPs listens to an instruction queue for measurement tasks (both VP-specific and network-specific instruction queues are supported), executes the tasks (if allowed according to the configuration of the VP), and returns results to the backend server via the message queueing system. Our current system prototype supports throughput-based measurements (NDT and wget), along with measurements such as traceroute, ping, and DNS lookups. The measurement system can be configured to restrict the set of measurements that a given VP may perform based on the capabilities and probing budget of that VP.

The reactive measurement system also allows us to specify additional measurement types, or to create complex measurements by composing the aforementioned measurement types. We believe that tests that measure or estimate the QoE of different application types would be valuable additions to this system. For example, in the recent network C-D example described in Section 3, it would have been useful to trigger additional experiments to estimate the QoE for applications running over that interdomain link, and whether the short-lived congestion episode had any perceptible impact on user QoE.

5. REFERENCES

[1] R. Andrews and S. Higginbotham. YouTube sucks on French ISP Free, and French regulators want to know why. *GigaOm*, 2013.

[2] J. Brodkin. Time Warner, net neutrality foes cry foul over Netflix Super HD demands, 2013.

[3] J. Brodkin. Why YouTube buffers: The secret deals that make-and break-online video. *Ars Technica*, July 2013.

[4] S. Buckley. France Telecom and Google entangled in peering fight. *Fierce Telecom*, 2013.

[5] Center for Applied Internet Data Analysis. Archipelago Measurement Infrastructure. `http://www.caida.org/projects/ark`.

[6] Y. Chen, S. Jain, V. K. Adhikari, and Z.-L. Zhang. Characterizing roles of front-end servers in end-to-end performance of dynamic content distribution. In *ACM SIGCOMM IMC*, Nov. 2011.

[7] S. Cowley. ISP spat blacks out Net connections. *InfoWorld*, 2005.

[8] J. Engebretson. Level 3/Comcast dispute revives eyeball vs. content debate, Nov. 2010.

[9] J. Engebretson. Behind the Level 3-Comcast peering settlement, July 2013. `http://www.telecompetitor.com/behind-the-level-3-comcast-peering-settlement/`.

[10] P. Faratin, D. Clark, S. Bauer, W. Lehr, P. Gilmore, and A. Berger. The growing complexity of Internet interconnection. *Communications and Strategies*, (72):51–71, 2008.

[11] M. Ricknas. Sprint-Cogent dispute puts small rip in fabric of Internet. *PCWorld*, Oct. 2008.

[12] Srikanth Sundaresan and Sam Burnett and Nick Feamster and Walter de Donato. BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks, 2014.

[13] Verizon. Unbalanced peering, and the real story behind the Verizon/Cogent dispute, June 2013. `http://publicpolicy.verizon.com/blog/`.