Understanding the Nature of Internet Congestion

Srikanth Sundaresan, Amogh Dhamdhere, kc Claffy, Mark Allman
International Computer Science Institute, Berkeley
Center for Applied Internet Data Analysis, San Diego

1 Abstract

The recent explosion of demand for high-bandwidth content (particularly streaming video) has led a renewed discussion about Internet congestion. Prior to the advent of large-scale video, there was an assumption in the academic, industrial, and the regulatory communities that bottlenecks were likely to be at the edge of the Internet path; *i.e.*, the access link or the home network. However, recent high-profile peering disputes between content, transit and access providers have resulted in the reevaluation of this assumption; bottlenecks occur not just in the edge, but also in the network core, in particular in locations where traffic transits from one network to another. Knowing where congestion occurs can be critical for applications and service providers to improve user quality of experience (QoE) for such applications. For example, knowing whether a video stream was congested by a peering link, or whether it was bottlenecked by a heretofore uncongested link (e.g., the user's access link) can be very useful in adapting to it—either by rerouting flows to avoid the congested peering link, or by reducing the bitrate to mitigate congestion on the access link. However, much of our understanding of such congestion events are based on circumstantial evidence and inconclusive measurements. In this body of work, we seek to enhance our understanding of the nature and location of Internet congestion and its implications on QoE; we propose to do so by analyzing and instrumenting bulk-transfer TCP flows to determine the nature and location of the bottleneck that limits them.

2 Why is understanding congestion important?

The rise of on-demand video has led to a sharp increase in the volume of Internet traffic. For instance, after Netflix started their video streaming service, their share of Internet traffic within the US rapidly rose, and currently accounts for about a third of the peak downstream traffic consumed by residential broadband users, according to some sources. This surge in demand for high-bandwidth content, and growing concentration of content among a few content distribution networks has led to an increasing number of incidents of congestion in "core" links; those inside and between large ISPs and transit providers. This is significant because it is no longer valid to assume that performance bottlenecks occur primarily in the last mile, as previous efforts were wont to do. For example, the FCC's Measuring Broadband America initiative focuses its measurements primarily on the last mile and the access ISP.

This development has several implications. The most important is on the ability of users and application providers to control QoE. The location and type of the bottleneck could determine the potential impact on user QoE, and guide the course of action needed to alleviate it. Congested peering links might spur the content provider to explore alternate transit or CDN schemes to deliver their content better. Access link bottlenecks might spur them to redesign the content itself so that they can deliver it more efficiently. The ability to accurately identify bottlenecks in a path could also avoid unedifying scenarios such as large ISPs, transit providers, and content providers pointing fingers at each other based on conjecture, without improving application performance. Similarly, this information can lead to users taking more control over their own QoE by making informed decisions about their choice of ISP, service plan, or content provider.

3 Research Agenda

The state of the art in our understanding of Internet congestion does not allow us to determine its cause or nature. We seek to change this in two broad ways: 1) developing techniques to identify whether a TCP flow was

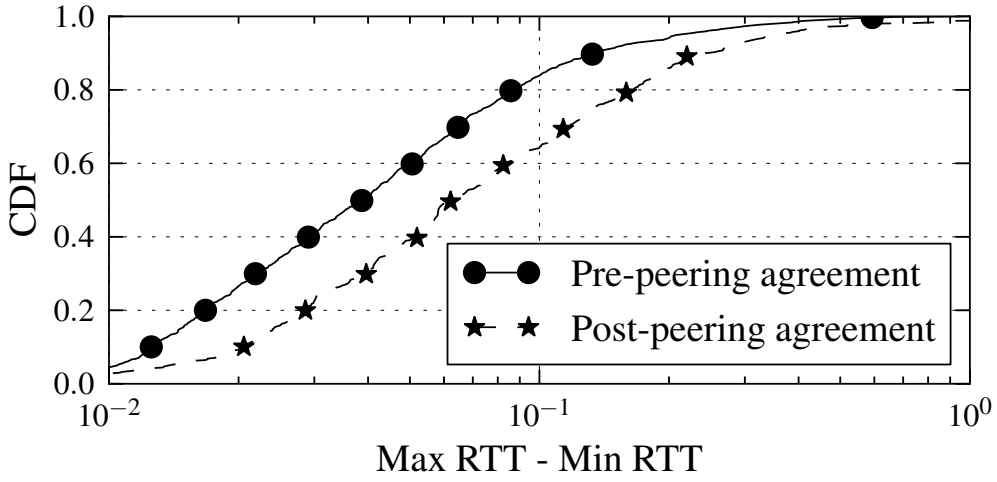

Figure 1: The difference between the maximum and minimum RTT during slow start. The difference is higher when TCP starts up on an uncongested path than when it starts up on an otherwise congested path.

bottlenecked because it increased its sending rate to fill a previously unconstrained link, or because it started up on an already congested link, and 2) developing techniques to localize the bottleneck.

- **Developing techniques to understand type of congestion:** We are working on a technique that allows us to understand whether a TCP flow was limited by an already congested link (e.g, a peering link in the core of the network) or by a link it filled up by scaling its sending rate (e.g, the user's access link). These two scenarios are qualitatively different; when a TCP flow starts up across a congested peering link, it must continuously push other competing flows out of the way to claim its share of bandwidth. On the other hand, when a TCP flow starts up on a hitherto unconstrained path, it experiences only momentary episodes of congestion as it probes for available bandwidth by adjusting its sending rate. These two scenarios could have very different implications for the QoE of applications and providers that use these paths: a congested peering link will require rerouting flows in the short run, and better peering, better content distribution in the long run, while a congested access link will require better content delivery mechanisms such as better better bitrate adaptation for video.

  Our intuition for differentiating these cases is based on TCPs slow start mechanism. When the path between the client and the server is not congested, TCP will increase its sending rate till it saturates one link in the path. This causes the buffer at the head of this link to fill up, in turn causing an increase in the end-to-end RTT of the flow. The RTT increases until the TCP flow experiences a loss and exits slow start. If the TCP flow re-enters slow start it will again start scaling up its sending rate, repeating the previous behavior. If the TCP flow enters congestion avoidance, the RTT will stabilize as the TCP flow probes for available capacity. In either case, the TCP flow itself is driving the queue, and therefore the RTT behavior. Conversely, when a link in the end-to-end path is already congested, the TCP flow must fight for a share of the pipe. It also always encounters a full buffer at the head of the congested link. This full buffer means not only that the baseline RTT is increased by the size of buffer, but also that the TCP flow does not experience the same kind of RTT increase over the course of the flow as experienced in the first case. In this case, flow RTT behavior is driven by aggregate flow behavior in the congested link. We are working on developing this intuition and build signatures based on TCP RTT samples to differentiate these two cases.

  Validating hypotheses about Internet congestion is difficult because of the scarcity of ground-truth data. Recently, however, data from Measurement Lab's NDT tests have offered an opportunity to study peering and last mile congestion events at large scale. We use this data to validate our earlier hypothesis about differentiating the nature of the bottleneck. We use data from a specific time period in 2014 when Netflix transitioned from using Cogent as transit to reach Comcast, to peering directly with Comcast. This caused a significant drop in congestion events seen between Comcast and Cogent, according to the 2014 Measurement Lab report on Interconnect congestion. We assume that the shift in Netflix traffic caused

NDT measurements between Comcast users and servers in Cogent to experience a shift in bottleneck from the core (Cogent/Comcast interconnect) to the edge (the access link). Simple TCP signatures based on RTT variation show that there is a significant difference in the variation of RTT during slow start for TCP flows collected before and after the peering agreement. The variation of RTT during slow start is higher in the period following the Comcast/Netflix peering agreement, which is what we would expect as those TCP flows were likely to be limited by the user's access link, and not by a congested peering link as flows prior to this period were likely to be. Figure 1 shows the CDF of the difference between the maximum and the minimum RTT for flows in these two periods: we see that there is a clear difference between the two, which validates our intuition. We are currently working on understanding and validating these signatures in a more robust manner.

- **Localizing the bottleneck:** While knowing that congestion was caused by a peering link is useful, it is important to be able to locate, at the link level (or at least at the AS level) where exactly the bottleneck occurred. We are working on building both in-band and out-band probing techniques that complement existing throughput tests to identify the bottleneck link. Current techniques use a time series of ttl-limited probes (the TSLP method from CAIDA/MIT). However they do not locate bottlenecks for live flows. We also plan to use TTL-limited probes, but also to pair those probes with throughput tests (or other flows). Out-band techniques, by which we mean outside of the flow we are trying to measure, are easier to instrument, but do not offer a guarantee that we are actually following the flow. In-band techniques are considerably harder to instrument without interfering with the flow, but offer a better chance of measuring the bottleneck that the flow experiences. In-band techniques also offer the ability for applications to detect the presence of throughput bottlenecks quickly, and estimate the possible impairments to QoE that such bottlenecks may cause. As the application itself is the best judge of QoE, quick feedback offers the application the ability to proactively take steps to maintain or improve QoE at fine-grained intervals.