Open access • Proceedings Article • DOI:10.1145/2663716.2663741

Challenges in Inferring Internet Interdomain Congestion — Source link

Matthew Luckie, Amogh Dhamdhere, David D. Clark, Bradley Huffaker ...+1 more authors

Institutions: University of California, San Diego, Massachusetts Institute of Technology

Published on: 05 Nov 2014 - Internet Measurement Conference

Topics: Internet topology, Internet traffic, Round-trip delay time and Latency (engineering)

Challenges in Inferring Internet Interdomain Congestion

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

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

David Clark
MIT
ddc@csail.mit.edu

Bradley Huffaker
CAIDA / UC San Diego
bradley@caida.org

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

ABSTRACT

We introduce and demonstrate the utility of a method to localize and quantify inter-domain 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, a delay pattern that differs from the typical diurnal pattern of an uncongested link. We describe our technique and its surprising potential, carefully analyze the biggest challenge with the methodology (interdomain router-level topology inference), describe other less severe challenges, and present initial results that are sufficiently promising to motivate further attention to overcoming the challenges.

Categories and Subject Descriptors

C.2.5 [**Local and Wide-Area Networks**]: Internet; C.2.1 [**Network Architecture and Design**]: Network topology

Keywords

Interdomain congestion; Internet topology

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.

IMC'14, November 5–7, 2014, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2663716.2663741.

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. [5,10, 32]) 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, 3, 4, 6, 13, 14, 37]), some large and sophisticated enough to adjust loading (and thus congestion levels) on interconnection links [9,15]. 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.

Unsurprisingly, there is growing public policy interest in the extent and scope of congestion induced by persistently unresolved peering disputes, and how harmful it is to consumers. Unfortunately, almost no data is available for researchers to study interconnection controversies. Traffic data and peering terms are almost always under NDA for newsworthy peering disputes; providers obfuscate network identities when they discuss congestion at all [35].

We provide three contributions to understanding the prevalence of interdomain congestion. First, we introduce and demonstrate the utility of a highly scalable probing method that allows us, from the edge of a given network to localize and characterize congestion on its interdomain links (section 2). Second, we analyze the many challenges associated with using this method to create a map of interdomain congestion, and how we have either started or plan to handle them (section 3). Third, we apply our method to illustrate evidence of persistent interdomain congestion involving large access and content providers (section 4). We compare our approach with related work in section 5 and identify ongoing future work in section 6.

2. TIME SEQUENCE 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.

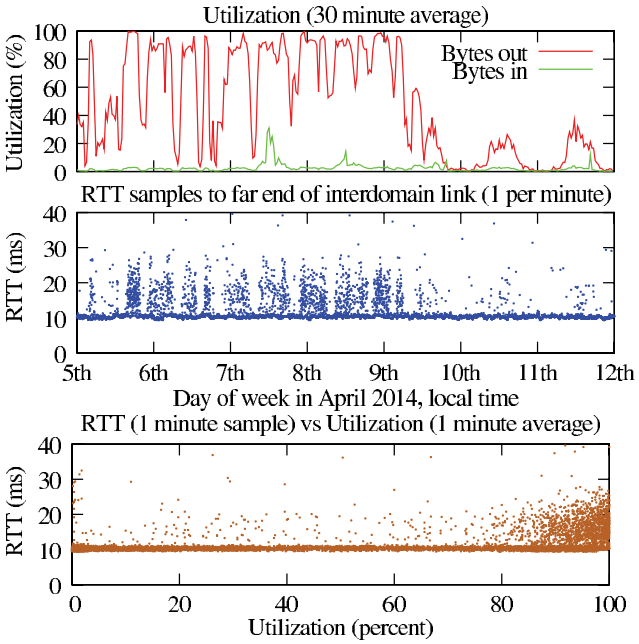

Figure 1: Comparing link utilization (top panel) with measured RTT (middle panel) on a 50Mbps customer link operated by a R&E network. The bottom scatterplot suggests that as the 1-minute average load exceeds ≈85%, increasing numbers of probes to the far side of the link encounter a queue, suggesting congestion at that moment.

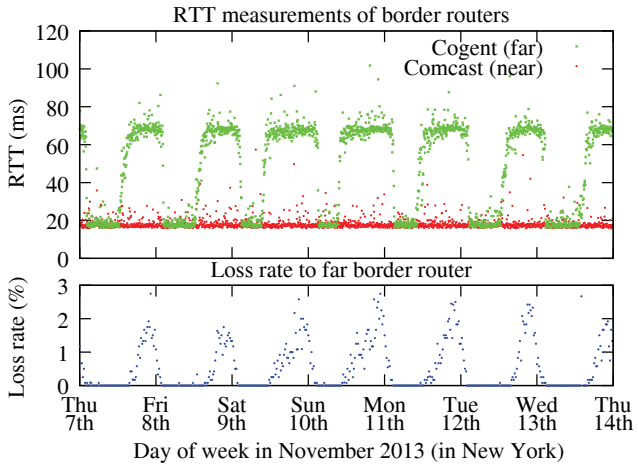

Figure 2: Congestion on an interdomain link between Comcast and Cogent, measured from a VP within Comcast. The RTT to the Cogent (far) router increases from 20ms to 70ms while the RTT to the Comcast (near) router is stable at 20ms. The approximate square wave indicates the queue is always close to full 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.

Lending some confidence to this method, figure 1 plots a week of traffic (SNMP byte counters sampled per-minute) and RTT measurements across a research and education network link known by the operator to be well utilized. The 30-minute average utilization on the link (top graph) correlates with periods when some probes experience increased RTT to the far end of the interdomain link (middle graph). The bottom graph shows that most RTT measurements above 10ms occur when the average utilization is above 85%. To maximize the chance of observing RTT variation across a specific link, TSLP sends TTL-limited packets toward the same destination that expire at the near and far routers, rather than send packets addressed to the border routers.

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 2 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.

We asked the ISPs involved to validate our inferences of this and other links that exhibit this behavior, but they are generally blocked by NDAs from sharing traffic data. Informal feedback from content, transit, and access network operators has given us confidence in our observations.

We believe the TSLP approach is a relatively lightweight method for obtaining data to map which interdomain links attached to the local network are congested. Compared with available bandwidth measurement techniques, such as pathload which sends 12 streams of 100 packets in 15 seconds [21] and requires tomography to identify which is the constraining link, TSLP uses 2 packets every 5 minutes to sample a targeted interdomain link, and does not require a cooperative end-host at the other end of the path. However, TSLP does send enough traffic that it would not scale to deployment in video players for diagnostic purposes. Many concurrent TSLP flows would trigger router ICMP response rate-limiting which defeats the method. Furthermore, TSLP requires a delay history to detect level shifts, and consumer video devices tend to operate only when the user wants to view a video. The value of TSLP is not in its potential universal deployment, but the insight that a remarkably sparse deployment can provide to all users sending or receiving traffic over TSLP-measured interconnection links.

Tulip [27] sends ICMP timestamp messages directly to routers to infer per-hop queuing delay for routers in the forward path as part of a system for diagnosing and pinpointing faults in Internet paths. Compared with their work, we are focused on finding which interdomain links are consistently underprovisioned, and we do not sample an interdomain link by sending packets directly to routers.

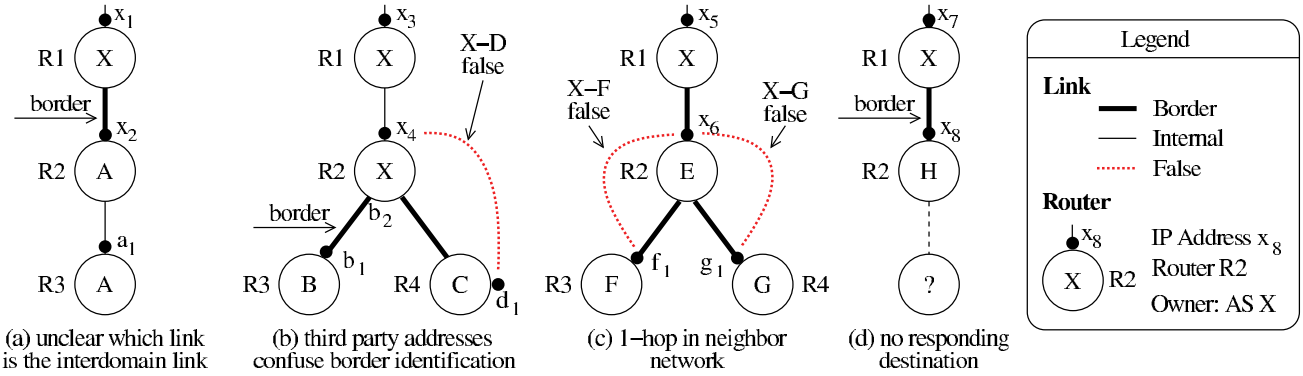

Figure 3: Challenges in AS boundary inference from traceroute for network X. In (a), it is unclear whether x_2, announced by X, or a_1 announced by A correspond to the far border router. In (b), X's peer C responds using an address d_1 originated by D, which could cause a false AS link inference X-D. In (c), we observe X's (unseen) peer with address x_6 originated by X, which could cause E's customers F and G to be incorrectly inferred as connected to X. In (d), we do not observe any responding address in H, which could cause us to not infer that x_8 represents H's border router.

3. METHODOLOGICAL CHALLENGES

While TSLP is a simple and surprisingly effective method for inferring congested links, there are many challenges to applying it effectively: identifying congestion on links with active queue management and/or weighted fair queueing policies; accurately finding and identifying all interdomain links involving the AS hosting a vantage point; proving the response from the far router returns over the targeted interdomain link; determining the direction of congestion; robustness to ICMP queuing behavior; adapting to path dynamics; and scaling processing to thousands of interdomain links.

3.1 AQM and WFQ

Active Queue Management (AQM) and Weighted Fair Queueing (WFQ) present challenges to TSLP. AQM techniques such as Random Early Detection [17] (RED) and successors try to maintain a small average queue size by discarding packets before a queue becomes full, as a function of the queue's current length. RED tries to prevent TCP global synchronization, where multiple senders experience packet loss at the same time and collectively reduce their transmission rate, resulting in an inefficient use of capacity. However, AQM techniques are not widely deployed [31], in part because operators traditionally provision links to meet forecasted demand; the only measurement study of RED deployment we know of was among access networks [12] rather than interdomain links. Where AQM techniques such as RED are deployed on congested links such as that in figure 2, we hypothesize that the level shift from low to high state will be more gradual as TCP connections encounter earlier loss, though we still expect to see the queue approach a state where it is nearly always full with this offered load.

A larger challenge is links that use a fair queueing strategy, where the router places packets in different queues according to some property, such as the incoming port or sender. TSLP cannot infer link congestion when our packets are placed in a queue that does not experience delay. From an interaction with one network operator, we learned that they had deployed WFQ on some of their persistently congested links; TSLP observed no level shift on these links, presumably due to the WFQ behavior.

3.2 Inferring Interdomain Links

If a public BGP view of the AS hosting a VP is provided to RouteViews or RIPE's RIS, then identifying interdomain links and paths where their border routers can be found is relatively simple. While some large access ISPs (including Verizon and AT&T) provide a public view, most large access ISPs (including Comcast, Time Warner, Cox, and RCN) do not, so we must infer AS links and corresponding border routers with traceroute, which is known to be difficult [8,39]. For each VP, we tracerouted to the first (.1) address in each of the 465,944 non-overlapping prefixes observed by RouteViews collector rv2, used BGP data to map the encountered IP addresses to ASes, and inferred an interdomain link when we observed the first address in a traceroute path that maps to an AS outside the VP's hosting network. However, this method may not correctly identify the border routers, or their owners, depending on how interfaces are numbered.

Figure 3 illustrates the variety of things that can go wrong. In figure 3a, when we observe traceroute path x_1, x_2, a_1, a simple IP-AS mapping incorrectly suggests the interdomain link is between R2 and R3, when it is between R1 and R2. More generally, the interdomain hop could be either the hop at which the IP-AS mapping changed, or one hop before. The convention in a customer-provider interconnection is to number the customer router interface from the provider's address space, which identifies R2 as the customer border router and the interdomain link as between R1 and R2, but there is no address assignment convention for peers.

It is usually simple to identify a border router connecting multiple peers as owned by the VP's network. However, some addresses observed after a border router may cause false interdomain link inferences because they map to *third-party* ASes, a well known challenge in AS topology mapping [39]. Figure 3b illustrates the danger: R4, owned by AS C, may respond with address d_1 which maps to AS D. We can improve the robustness of our inferences in the presence of third party addresses with two heuristics. First, if we require interface d_1 to have been observed in a traceroute path toward a prefix announced by D, d_1 is unlikely to be a third-party address. However, this filter discards many true adjacencies if no paths toward a prefix in B cross a specific interconnection with B. A second heuristic can re-

tain some of these adjacencies by proving b_1 is not a third-party address: b_1 represents the incoming interface on R3 and a point-to-point interdomain link between X and B if b_2 is in the same /30 or /31 subnet and is an alias of x_4. We used this heuristic to prove that most addresses in a traceroute path represent the inbound interface on a point-to-point link [26]. Finally, if the interfaces used to infer an adjacency with D were x_4 and d_1, and address d_1 was probed, we have learned to be skeptical of the x_4 - d_1 adjacency, as a router with d_1 configured will reply to probes to d_1 with that address; d_1 might be configured on R3 and connect a subnet one hop away. When we receive a response directly from d_1, we try to discern the router it is connected to by probing another address in the same prefix as d_1 to solicit a TTL-exceeded response from the router on the path.

We have encountered related cases where we only observed the neighbor's border router and no further hops owned by the neighbor network. Figure 3c illustrates the implications of this challenge: R2 is owned by E, but we observe only x_6 on the router, and subsequent interfaces f_1 and g_1 could falsely imply interdomain links between X and F, G. We resolve these ambiguities by finding a *common provider AS* E for both F and G which we derive from CAIDA's AS relationship inferences [25].

In some cases we never observed an address in a neighbor network. Figure 3d illustrates this challenge: R2 is owned by H but R2's address x_8 is the last address we observed in the traceroute. To infer R2 is owned by H, we infer a *common origin AS* to the prefixes probed.

Chen *et al.* [8] used traceroute and BGP data to derive AS links from traceroute paths using heuristics more robust than a simple IP-AS mapping. Their work addressed the third-party address (figure 3b) and one-hop in neighbor network (figure 3c) problems by comparing AS paths inferred with traceroute to AS paths observed in BGP for the corresponding prefix. If they found an extra AS hop (D in X-D-C in figure 3b) they removed D from the traceroute-inferred AS path. If they found a missing AS hop (E in X-E-F in figure 3c) they added E to the traceroute-inferred path. However, they do not adjust IP-AS mappings or assign owners to routers. Integration of their techniques would likely improve our router ownership heuristics.

Mao *et al.* [29] adjusted IP-AS mappings by changing the owner of /24 prefixes to make traceroute-inferred AS paths congruent with BGP AS paths collected in the same AS. However, individual addresses (x_2 and x_8 in figure 3) are not mapped to different owner ASes. Zhang *et al.* [38] adjusted IP-AS mappings using the same approach as Mao *et al.* [29] but at /32 granularity; they did not use AS relationship data (customer relationships) to infer owners of border routers so did not adjust x_2 unless it made traceroute AS paths more congruent. Huffaker *et al.* [19] evaluated heuristics using router alias, AS relationships, and AS degrees. For routers with addresses from multiple ASes, assigning ownership to the AS with most addresses on the router yielded the most accurate results.

3.3 Asymmetric paths

A TTL-limited response from the far border router might not return via the near router, because the far router is operated by a different AS that might have a more preferred path toward our VP. In general, we hypothesize that a peer will respond via the near router, a provider will respond via the near router except in cases where the VP's AS (i.e., the customer) is doing traffic engineering, and a customer will respond via the near router unless it has a lower-cost path. We evaluated two methods to gain confidence that an increase in measured RTT from a near to a far router is due solely to traffic on the link connecting the two routers; that is, the link behavior was *isolated*. Katz-Bassett *et al.* used these two methods in reverse traceroute [22].

Pre-specified timestamps (PSTS): The PSTS IP option allows a host to request other hosts embed a timestamp in the packet. Using the notation in [33], G|BCDE denotes a probe to destination G that requests hosts with addresses B, C, D, and E to include timestamps. The option includes a pointer to the next timestamp slot in the packet, which advances when a router embeds a timestamp; if the packet visits C but not B first, then C will not embed a timestamp. Using the topology in figure 3a, to test if R3 returns packets to our VP via R2, we send an ICMP echo request packet in the form $a_1|a_1x_2$. If both a_1 and x_2 embed a timestamp, we infer the packet was returned across the R2-R3 link.

Record Route (RR): The RR IP option allows a host to request that up to nine hosts embed an IP address as they forward the packet. To test if R3 returns packets to our VP via R2, we send an ICMP echo request packet with the RR option set to R3. If we observe an address belonging to R3 in the RR option in the response, and an address belonging to R2 immediately after, we infer the packet returned across the R2-R3 link.

Preliminary Evaluation: Both the PSTS and RR options are known to have limitations due to routers that either do not implement the functionality, discard packets that contain options, or (in the case of RR) do not have sufficient space to embed the addresses of interest. Of the 599 interdomain router links involving Comcast that we assembled from our Ark VP (mry-us) deployed in Monterey, California we use either PSTS or RR to infer that 179 (29.9%) returned over the targeted link; 72 (12.0%) were isolated only with RR, and 78 (13.0%) were isolated with only PSTS. We could not isolate 71 (11.9%) because all nine slots were used, and the remainder (58.2%) were unresponsive to these IP options or inconclusive; different addresses in traceroute and RR may belong to the same router but resist alias resolution. We manually checked a few paths that RR suggests did not return over the targeted link; we observed, for example, Comcast's provider Tata forwarding packets that it received in Los Angeles to the San Jose interconnection.

3.4 Other challenges

ICMP queuing behavior: A concern with using ICMP TTL-exceeded responses is that routers may delay these responses (process them through the slow path), especially during periods of high load; we may thus measure load on the router and not congestion on a specific link. But it is unlikely that slow path processing would induce the same delay for each probe response, so using the minimum RTT during a given time window will more likely reflect the queue size, improving TSLP's robustness to potential idiosyncrasies in ICMP behavior. In some routers, the ICMP response generation delay spikes every 60 seconds due to periodic maintenance activity [27]. This behavior will mislead TSLP when our 5-minute samples synchronize with the maintenance activity. To avoid this problem we could randomize the send time of our probes in each 5-minute measurement round.

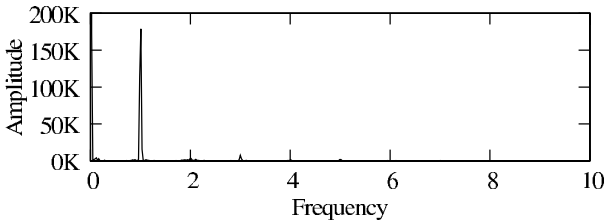

Figure 4: FFT analysis of time series data from the Comcast-Cogent interdomain link in figure 2. The sharp peak at f=1/day implies a strong diurnal pattern in the time series.

Determining the direction of congestion: Even when we are confident that the TTL-exceeded response from the far end of the router returns via the target link, we do not know whether the congestion on that link is in the forward or reverse direction (from the perspective of our VP). To determine the direction of congestion on the targeted link we could use the prespecified timestamp option, sending a sequence of probes toward the target link, soliciting timestamp t_n^1 from the near router, t_f from the far interface, and t_n^2 from the near router again. Clock skew between the routers prevents use of the difference in these timestamps to estimate queuing delay; however, if t_f - t_n^1 shows a diurnal pattern, then we can infer that the link is congested in the forward direction. If t_n^2 - t_f shows a diurnal pattern, then the link is congested in the reverse direction.

Adapting to change: Our probing setup infers interdomain links from traceroutes, and notes the distance from the VP at which each target is seen. However, network routing may change over time; the path to a destination may traverse a different interdomain link, or the same interdomain link may be seen at a a different distance from the VP. To adapt to change, each VP runs a topology discovery process in the background to continuously map interdomain links and their distance from the VP. We then adapt our probing to respond to changes in the measured topology.

Automated trace processing: A VP that tests every interconnection link out of its AS can yield hundreds (access) or thousands (tier-1) of interdomain links, requiring some automated method to detect evidence of congestion. In our initial study, a repeating diurnal pattern with a consistent duration of RTT change, such as that presented in the Comcast-Cogent interdomain link in figure 2, manifested clearly in a frequency domain transformation using a Fast Fourier Transform (FFT) with power density at f = 1/day in figure 4. A time series of RTT samples that contains a power density at 1 can be automatically identified as interdomain congestion. However, some RTT patterns imply congestion is present for only part of the week or weekend. In this case a wavelet transform may reveal the structure of frequencies across time. In the limit, there is a decision as to whether these cases represent noteworthy congestion, which is not an issue any classifier will resolve.

4. CASE STUDIES

We present five case studies showing the potential of the TSLP method to provide empirical data on peering disputes.

Evolution of congestion on Comcast links: A residential Ark monitor in Comcast's network in Monterey, California continuously performs ICMP Paris traceroutes toward randomly chosen destinations in all routed IPv4 /24 prefixes [20]. Large transit networks that are in the path to many destinations have their interdomain links sampled frequently. Our packets traversed interdomain links with Cogent, Level3, and Tata, which also transited Netflix traffic. To discover elevated RTTs in this data we used a level shift detection method that relies on the rank-based nonparametric statistical test CUSUM [36], which is robust to outliers and makes no assumptions about the distribution of underlying RTTs. Figure 5 shows the inferred duration of congestion on these links (in hours per day) from February 2013 to April 2014. Both the Cogent and Level3 links grow from being congested 4-6 hours per day in February 2013 to a maximum of 18 hours (Cogent) and 14 hours (Level3) in February 2014. From mid-May to mid-July, the congestion signal on the Level3 link is replaced with congestion on the TATA link, suggesting a significant volume of traffic was shifted from the Level3 to the TATA link. In late February 2014, Netflix and Comcast agreed to peer directly, and then congestion on the Cogent and Level3 links disappeared.

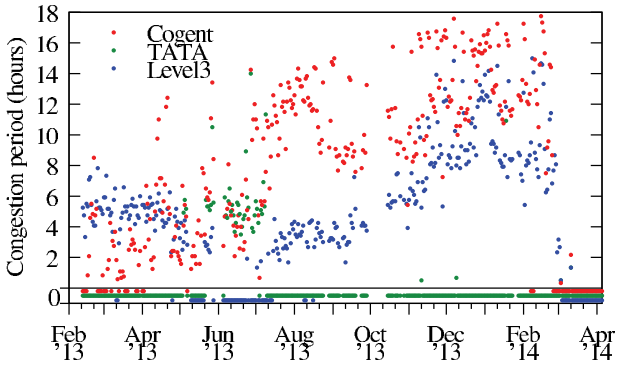

Figure 5: Inferred congestion duration (using CUSUM-based [36] level shift detection) for links connecting three major networks to Comcast. By February 2014, the Cogent and Level3 links were congested up to 18 and 14 hours per day, respectively. After Netflix and Comcast signed a peering agreement in February 2014, congestion on those links disappeared.

Netflix and Comcast direct peering: After the February 2014 agreement between Comcast and Netflix, our TSLP probes from the Comcast network started traversing direct peering links toward Netflix prefixes. For most interconnections, there was no level shift in RTT values that indicated a queue was always close to full. However, the peering link between Comcast and Netflix in San Jose, CA still appeared congested for 2-3 hours per day in April 2014 (figure 6a). We asked Netflix about it and learned that they had not fully deployed their peering with Comcast at San Jose.

Google and Free: Inspired by customer reports of poor performance of Youtube videos on Free's network [1], we used our Ark monitor in Free's network to measure the near and far end of a link between Google and Free with TSLP. Figure 6b shows a link that appears congested for 4-6 hours at peak time on weekdays, and more than 12 hours on the weekends (March 22nd and 23rd).

Level3 and AT&T, Verizon: In April 2014, Level3 published an article on their persistently congested links with six large broadband consumer networks: five were in the US and one in Europe [35]. They published an MRTG graph of their interconnect with an unnamed peer in Dallas,

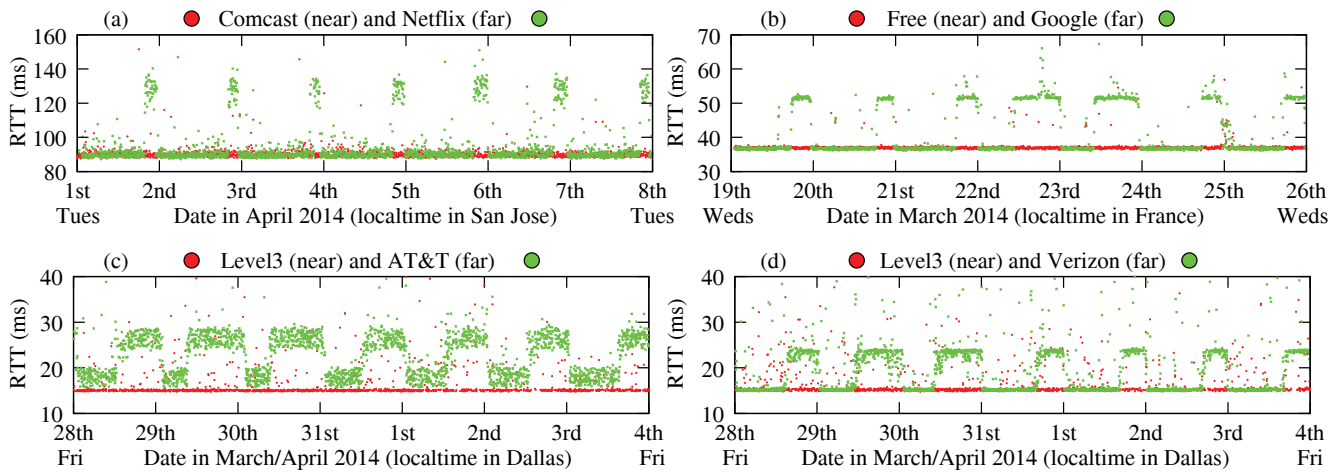

Figure 6: Case studies of four interdomain connections, discussed in section 4.

Texas. From our Ark monitor in Level3's network, we were able to observe congestion on Level3's interconnections with both AT&T and Verizon (we presume two of the broadband consumer networks) in Dallas (figures 6c and 6d). These links appear congested for 6 hours during weekdays and up to 16 hours during weekends in March/April 2014, and both are consistent with the anonymized MRTG graph published by Level3. While both time series are similar for this selected week (we chose to coincide with the MRTG graph in the Level3 article) we observe differences in the diurnal pattern on March 31st: AT&T was congested longer, and we believe it is the anonymized example [35]. For Verizon, the last day we observed the diurnal pattern was June 22nd. For AT&T, the last day we observed the pattern was July 14th; our Ark monitor was down until the 19th, and the signal was gone when the monitor restarted. This example illustrates the power of our technique to (1) provide independent evidence of specific congestion, (2) identify the parties involved where the parties themselves cannot do so because of their NDA agreements, and (3) inform the public debate.

5. OTHER RELATED WORK

Current methods of measuring systematic congestion require instrumentation at both ends of a path being measured [18, 30], but our TSLP method can potentially measure congestion from any VP hosted in an AS to any interdomain link involving that AS provided the link can be isolated. More recently researchers have studied in-home [7,11], and broadband access [2, 16, 23, 34] performance issues, and whether these components are the end-to-end performance bottlenecks for most users. Liu and Crovella used simulation to show that the *loss-pairs* methodology [24] can infer if a router uses AQM or tail-drop. Mahimkar *et al.* used a level shift detection algorithm to correlate changes in network performance with network upgrades [28].

6. CONCLUSION

We developed and tested a simple method of identifying congestion on interdomain links, and used it to study several incidents of reported congestion on interdomain links that correlate with reports of contested business negotiations between the ASes. The advantages of this method are its conceptual, implementation, and deployment simplicity. In contrast to experiments that produce broadband performance maps, which require VPs at many access points, we can measure interdomain links from a given serving area of an ISP with one VP. Most importantly, the TSLP method does not require a server on the other side of the link being probed. Approaches that depend on a server may reveal evidence of congestion from one measurement, but require either instrumentation in the right two places, and/or complex correlation and tomography to localize the point of congestion. Since the vast majority of links do not exhibit persistent congestion, being able to localize congestion deep in the network from a single endpoint has benefits that justify further attention by the research community to resolve the many challenges we have described.

The major challenge is not finding evidence of congestion but associating it reliably with a particular link. This difficulty is due to inconsistent interface numbering conventions, and the fact that a router may have (and report in ICMP responses) IP interface addresses that come from third-party ASes. This problem is well understood, but not deeply studied. Going forward we will study this problem, as well as focus on localizing the directionality of congestion.

We have reported early results from a new project to detect and measure congestion at Internet interconnection points, an issue of recent interest due to the changing nature of Internet traffic over the last decade. For the thousands of interconnection links we have probed thus far, we have not found evidence of widespread persistent congestion, which is good news, but also suggests the value of a lightweight technique that can locate interdomain congestion from a VP at the edge. We plan to deploy VPs in as many access networks as possible, to generate a global "congestion heat map" of the Internet, cross-validate with non-ICMP traffic, and publish evidence of congestion over time.

Acknowledgments

We thank the operators who discussed aspects of their network's operations, Ahmed Elmokashfi who provided an implementation of CUSUM-based level shift detection, and the anonymous reviewers and Ethan Katz-Bassett for their feedback. This work was supported by the U.S. NSF CNS-1414177 and CNS-1413905, by Comcast, and by Verisign, but the material represents only the position of the authors.

7. REFERENCES

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

[2] S. Bauer, D. Clark, and W. Lehr. Understanding Broadband Speed Measurements. In *TPRC*, 2010.

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

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

[5] S. Buckley. Cogent and Orange France fight over interconnection issues. *Fierce Telecom*, 2011.

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

[7] K. L. Calvert, W. K. Edwards, N. Feamster, R. E. Grinter, Y. Deng, and X. Zhou. Instrumenting home networks. *ACM SIGCOMM CCR*, 41(1), Jan. 2011.

[8] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao. Where the sidewalk ends: Extending the Internet AS graph using traceroutes from P2P users. In *ACM CoNEXT*, Dec. 2009.

[9] 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.

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

[11] L. DiCioccio, R. Teixeira, M. May, and C. Kreibich. Probe and Pray: Using UPnP for Home Network Measurements. In *PAM*, pages 96–105, 2012.

[12] M. Dischinger, A. Haeberlen, K. P. Gummadi, and S. Saroiu. Characterizing residential broadband networks. In *ACM SIGCOMM IMC*, Oct. 2007.

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

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

[15] 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.

[16] FCC. Measuring Broadband America, 2011. `http://www.fcc.gov/measuring-broadband-america`.

[17] S. Floyd and V. Jacobson. Random early detection (RED) gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, 1993.

[18] D. Genin and J. Splett. Where in the Internet is congestion?, 2013. `http://arxiv.org/abs/1307.3696`.

[19] B. Huffaker, A. Dhamdhere, M. Fomenkov, and kc claffy. Toward topology dualism: Improving the accuracy of AS annotations for routers. In *PAM*, Apr. 2010.

[20] Y. Hyun, B. Huffaker, D. Andersen, M. Luckie, and K. C. Claffy. The IPv4 Routed /24 Topology Dataset, 2014. `http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml`.

[21] M. Jain and C. Dovrolis. End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput. *IEEE/ACM Transactions on Networking*, 11(4):537–549, 2003.

[22] E. Katz-Bassett, H. V. Madhyastha, V. K. Adhikari, C. Scott, J. Sherry, P. vas Wesep, T. Anderson, and A. Krishnamurthy. Reverse traceroute. In *USENIX NSDI*, Apr. 2010.

[23] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: Illuminating the Edge Network. In *ACM SIGCOMM IMC*, 2010.

[24] J. Liu and M. Crovella. Using loss pairs to discover network properties. In *ACM SIGCOMM IMW*, Nov. 2001.

[25] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and k claffy. AS relationships, customer cones, and validation. In *ACM SIGCOMM IMC*, 2013.

[26] M. Luckie and kc claffy. A second look at detecting third-party addresses in traceroute traces with the IP timestamp option. In *PAM*, 2014.

[27] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson. User-level Internet path diagnosis. In *ACM SOSP*, Oct. 2003.

[28] A. A. Mahimkar, H. H. Song, Z. Ge, A. Shaikh, J. Wang, J. Yates, Y. Zhang, and J. Emmons. Detecting the performance impact of upgrades in large operational networks. In *ACM SIGCOMM*, Aug. 2010.

[29] Z. M. Mao, D. Johnson, J. Rexford, J. Wang, and R. Katz. Scalable and accurate identification of AS-Level forwarding paths. In *IEEE INFOCOM*, Mar. 2004.

[30] M. Mathis. Model Based Bulk Performance Metrics. `http://datatracker.ietf.org/doc/draft-ietf-ippm-model-based-metrics/`, Oct 2013.

[31] K. Nichols and V. Jacobson. Controlling queue delay. *ACM Queue*, 10(5), May 2012.

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

[33] J. Sherry, E. Katz-Bassett, M. Pimenova, H. V. Madhyastha, T. Anderson, and A. Krishnamurthy. Resolving IP aliases with prespecified timestamps. In *ACM SIGCOMM IMC*, Nov. 2010.

[34] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè. Broadband Internet performance: A view from the gateway. In *ACM SIGCOMM*, 2011.

[35] M. Taylor. Observations of an Internet Middleman, May 2014. `http://blog.level3.com/global-connectivity/observations-internet-middleman/`.

[36] W. A. Taylor. Change-point analysis: A powerful new tool for detecting changes, 2000. `http://www.variation.com/cpa/tech/changepoint.html`.

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

[38] B. Zhang, J. Bi, Y. Wang, Y. Zhang, and J. Wu. Refining IP-to-AS mappings for AS-level traceroute. In *IEEE ICCCN*, July 2013.

[39] Y. Zhang, R. Oliveira, H. Zhang, and L. Zhang. Quantifying the pitfalls of traceroute in AS connectivity inference. In *PAM*, 2010.