Challenges in Inferring Internet Congestion Using Throughput Measurements

Srikanth Sundaresan
Princeton University
srikanths@princeton.edu

Danny Lee
Georgia Tech
dannylee@gatech.edu

Xiaohong Deng, Yun Feng
University of New South Wales
{xiaohong.deng,yun.feng}@unsw.edu.au

Amogh Dhamdhere
CAIDA, Univ. of California, San Diego
amogh@caida.org

ABSTRACT

We revisit the use of crowdsourced throughput measurements to infer and localize congestion on end-to-end paths, with particular focus on points of interconnections between ISPs. We analyze three challenges with this approach. First, accurately identifying which link on the path is congested requires fine-grained network tomography techniques not supported by existing throughput measurement platforms. Coarse-grained network tomography can perform this link identification under certain topological conditions, but we show that these conditions do not always hold on the global Internet. Second, existing measurement platforms provide limited visibility of paths to popular web content sources, and only capture a small fraction of interconnections between ISPs. Third, crowdsourcing measurements inherently risks sample bias: using measurements from volunteers across the Internet leads to uneven distribution of samples across time of day, access link speeds, and home network conditions. Finally, it is not clear how large a drop in throughput to interpret as evidence of congestion. We investigate these challenges in detail, and offer guidelines for deployment of measurement infrastructure, strategies, and technologies that can address empirical gaps in our understanding of congestion on the Internet.

CCS CONCEPTS

• **Networks** → **Network measurement**;

KEYWORDS

Internet congestion, Internet topology, Throughput

ACM Reference Format:
Srikanth Sundaresan, Danny Lee, Xiaohong Deng, Yun Feng, and Amogh Dhamdhere. 2017. Challenges in Inferring Internet Congestion Using Throughput Measurements. In *Proceedings of IMC '17, London, United Kingdom, November 1–3, 2017,* 14 pages.
https://doi.org/10.1145/3131365.3131382

IMC '17, November 1–3, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5118-8/17/11…$15.00
https://doi.org/10.1145/3131365.3131382

1 INTRODUCTION

The relentless growth of Internet traffic demands, and the growing concentration of content across a few providers and distribution networks, have led to capacity constraints, particularly at points of interconnection between content providers, transit providers, and access ISPs. Interconnection bandwidth contention has in turn led to high-profile disputes over who should pay for additional interconnection capacity [8–10, 19, 40, 43]. The resulting and potentially contentious interactions among providers have implications for network stability and performance, leaving the congested link as an externality for all users of the link until the dispute is resolved. This situation has led to recent interest in technical, regulatory, and policy circles in techniques to detect the presence of persistent interdomain congestion, and to localize it to specific links in the network. One approach to detecting congestion that has received significant recent attention is the use of crowdsourced throughput measurements, such as those offered by the Measurement-Lab (M-Lab) platform or Ookla's Speedtest.net web service [31].

We present a systematic analysis of inference techniques and challenges associated with using throughput measurements to infer the presence, location, and characteristics of congestion. We use NDT tests collected by the M-Lab platform in May 2015, along with M-Lab's analysis of this data [4, 27], as a case study of throughput-based inferences, and explore three challenges.[1]

First, using crowdsourced end-to-end throughput measurements to localize congestion to specific links in the network requires applying tomographic techniques to network paths observed during the measurements. We show that applying tomography at a coarse AS-level is difficult, requires several assumptions about the topology that do not always hold, and can be complicated by the complexity of link and router-level interconnection that constitute an AS-link.

Second, platforms capable of supporting throughput measurements currently observe a limited set of interdomain links of any given access network. While M-Lab operates a large distributed server-side measurement infrastructure and releases all data publicly, as of February 2017 M-Lab was able to measure between 0.4% and 9% of AS-level interconnections of access ISPs in U.S. (between 2.8% and 30% when considering AS-level peer interconnections). Furthermore, we show that between 79% to 90% of AS-level interconnections traversed on paths from U.S. ISPs toward popular web content were not testable using M-Lab's server infrastructure.

[1] We used data from 2015 to align it with the M-Lab reports [4, 27].

Finally, although crowdsourcing measurements can be an excellent way to expand sampling coverage and diversity, crowdsourcing can also yield biased samples. Coverage by user-generated measurements around the world does not account for other factors that may influence performance, including time of day, access link speed, and quality of the home network launching the tests. The opacity of these factors prevents reasoning about their possible influence, which raises concerns about the statistical validity of the analyses.

After a systematic analysis of available data, we offer several suggestions for improving the utility of collected throughput measurements to characterize congestion upstream of a client's access link. These suggestions include improved topology measurement and analysis techniques, more careful stratification of test results based on topological information, strategic deployment of server infrastructure to maximize coverage, and cross-validating crowdsourced measurements using more systematically collected data from other measurement platforms. Our analysis informs our own planning of existing and future measurement infrastructure, not just for throughput measurement, but for any generalized framework that aims to measure performance on an Internet-wide scale.

Section 2 presents details of existing measurement infrastructure projects that support crowdsourced throughput measurements and publish results. Section 3 and 4 describe how to apply tomographic techniques to these measurements to infer the location of congestion, and explain limitations of these techniques. Section 5 describes limitations in visibility of relevant interconnections, and Section 6 reviews statistical limitations on use of crowdsourced throughput measurements. Section 7 summarizes lessons learned and offers recommendations for improving interconnection congestion measurement and inference capability.

2 EXISTING THROUGHPUT MEASUREMENT PLATFORMS

Multiple platforms solicit crowdsourced throughput measurements from users using Web-based speed tests, including Ookla's Speedtest.net, DSLreports, and M-Lab. In addition, clients that are part of the FCC's Measuring Broadband America [44] platform or the Bismark platform [7] also perform periodic throughput measurements from home routers. Throughput measurements typically run simple bulk data transfers over TCP [31] over a short period of time, aiming to measure the bottleneck link by saturating it with one or more TCP flows. The bottleneck link is commonly the "last mile": the access link between the client and the Internet. In this scenario, the ideal location of the server is as close as possible to the client, to minimize latency to the client. TCP throughput has a well-understood inverse relationship with latency [33]: the longer the latency across a path, the lower the throughput, all other factors being equal. Therefore, as broadband access speeds increase, low latencies from test servers to clients ensure that throughput measurements can saturate the bottleneck link on the path to the client[6]. Popular throughput measurement service providers such as Ookla and M-Lab have extensive geographically distributed infrastructure to achieve low latencies to clients.

2.1 Throughput Measurements on M-Lab

M-Lab is a distributed platform with hundreds of well-provisioned machines around the world that serve as destinations (targets) for a set of freely available performance measurement and diagnostic tools. Users may download and run any of the supported software tools that estimate performance characteristics of paths between the client and M-Lab servers. One of these tools, the Network Diagnostic Test (NDT), is a Web-based tool that runs a throughput measurement in each direction: from client to server (upstream), and server to client (downstream), as follows. A client initiates an NDT measurement to M-Lab, and the M-Lab backend uses IP geolocation to select a server close to the client. Each test estimates the *downstream throughput* from the server to the client, and the *upstream throughput* from the client to the server, during which the server logs statistics including round trip time, bytes sent, received, and acknowledged, congestion window size, and the number of congestion signals (multiplicative downward congestion window adjustments) received by the TCP sender. The server also stores the raw packet captures for the test, which are publicly available through Google's BigQuery and Cloud Storage [29, 30]. Unlike Ookla and DSLreports, which only make aggregated stats available publicly, M-Lab makes all data available, including packet traces and supplementary path data.

2.2 Inferring Congestion using M-Lab data

The high density of U.S. server deployments on M-Lab facilitates the crowdsourcing of a rich set of measurements that include different combinations of transit provider and access ISPs. In 2014 and 2015, two well-publicized measurement reports—one from M-Lab itself [27], and another from the "Battle for the Net" advocacy group [1]—used NDT measurements on M-Lab to infer congestion in peering and transit networks in the U.S.

In 2014, a team from M-Lab team analyzed 18 months of NDT data collected by the M-Lab platform to infer interdomain congestion between large transit ISPs and large access ISPs [27]. The report aggregated results from clients of access ISPs to servers located in various transit providers, and grouped the tests by source AS, destination AS, and server location. They analyzed metrics such as download throughput, flow round-trip time (or flow RTT, which is the latency between the server and the client), and packet retransmission rates. They found diurnal patterns in the median values of these metrics, from which they inferred persistent congestion on paths between several U.S. access ISPs (including many large ISPs such as Comcast, AT&T, Verizon, and Time Warner) and transit providers (such as Cogent, Verizon, and XO). This report generated discussion in the academic community regarding the technical soundness of the measurement and analysis methods, and requests for revisions to the report to address its flaws [41].

Starting early 2015, a net neutrality advocacy group called "Battle for the Net"[1] hosted a modified version of the NDT client on their website. Their client was essentially a wrapper around NDT which performed back-to-back tests with up to five M-Lab servers in the same region rather than just the closest one, in an attempt to observe more paths. This group released, and then significantly modified [42], a report claiming evidence of congestion in more networks than the original M-Lab report. Media coverage generated a

jump in the number of NDT tests across M-Lab in May 2015, which prompted an M-Lab researcher to issue an update to M-Lab's congestion report in a blog post [4]. The posting described how they applied the same inference method (from [27]) on the newer data set to infer evidence of congestion in more interconnection links (such as from Verizon, Comcast, and Time Warner to GTT, and TATA). In the meantime, a public comment in the policy debate on the AT&T-DirectTV merger approved in 2015 cited the 2014 M-Lab report [27] as justification to propose severe regulatory conditions on AT&T following the merger including mandated settlement-free peering with some peers, and mandated upgrades to interconnection links upon reaching 70% utilization [15]. This public comment to the FCC illustrated the need for objective, peer-reviewed analysis of the state of interdomain congestion measurement methodology. In this paper we offer an attempt at such an analysis.

In June 2016, Google incorporated speed tests (driven by NDT) into Google search, allowing a sample of users that visited the Google search page to execute NDT tests against M-Lab servers [17]. By December 2016, the number of monthly NDT tests had increased more than 4-fold as compared to June 2016. After fixing some issues with the Paris traceroute data collection [35], the M-Lab platform now collects a large volume of NDT measurements along with Paris traceroutes from the server to client. The widespread interest in the previous analyses of the M-Lab data in technical and policy circles, along with the increased volume of NDT data in recent months represents a tempting opportunity to repeat the analysis of 2014-2015 in recent months.

We emphasize that our goal in this work is not to challenge the conclusions presented in the M-Lab reports, but instead to highlight the pitfalls and challenges in inferring congestion using throughput-based tests, and to illustrate the use of path measurements paired with throughput tests for more rigorous analysis. We hope that doing so will encourage improvements in the testing platforms to better support inference and localization of congestion, and analysis that more rigorously accounts for the complexity of interdomain interconnection. We use the M-Lab data and the previously mentioned analyses [4, 27] as case studies, because they were the first to attempt to use throughput measurements to infer and localize interdomain congestion. As such, the presentation in this paper necessarily delves into the particulars of the M-Lab infrastructure and analysis. Nonetheless, we believe that the results and recommendations are applicable to other platforms that attempt such analysis.

3 USING NETWORK TOMOGRAPHY TO INFER CONGESTION

Given a set of end-to-end measurements of some metric of interest (such as throughput, delay, loss rate or reachability) and knowledge (or measurements) of topology, one can use network tomography to infer the properties of each link in the topology. Binary network tomography [18] constrains the tomography problem by assuming that network-internal links can be in one of two states — "good" or "bad", and then attempts to find the smallest set of "bad" links that are consistent with the end-to-end observations. In the context of congestion localization, the end-to-end metric is an estimate of whether the path is congested, and links in the network can either be "congested" or "not congested".

Unfortunately, path information is not always available from large-scale throughput measurement platforms. M-Lab collects *Paris traceroutes* [5] from M-Lab servers toward clients that run measurements against their infrastructure, and releases this data publicly. Paris traceroute overcomes the problems with using the traditional traceroute tool for inferring router-level topology and paths in the presence of load-balancing, by carefully controlling the header fields in sent packets. However, the path information from Paris traceroute was incomplete prior to 2015 (Section 4.1) and in the latter half of 2016 [35]. Other large-scale throughput measurement platforms such as Ookla's Speedtest.net either do not collect path information at all, or do not release this data. Furthermore, even if path information is available from traceroutes, using it as input to a tomography algorithm is challenging due to issues with measurement synchronization and traceroute artifacts [21]. One alternative is to use a simplified form of tomography at the AS-level; M-Lab's studies of interconnection congestion used this simplified approach. This section describes this method and its assumptions.

3.1 Applying simplified AS-level tomography for congestion localization

Strong diurnal trends in achieved throughput in NDT measurements from servers in an AS S (source AS) to clients in an AS A (access AS) indicate the presence of link(s) along the path between S and A that are congested at peak times. However, other problems unrelated to congestion could result in such diurnal trends in NDT download throughput, e.g., problems in the user's home network or the user's access network. One way to mitigate the effects of access link issues is to compare diurnal congestion trends seen in NDT measurements from an access network A to servers in different ASes. If paths from a source AS S_1 to access network A show diurnal patterns indicating peak-hour congestion, but paths from source AS S_2 to access AS A do not, this difference suggests that access link/home network congestion was not a factor. The M-Lab reports went further and claimed that *any* perceived performance degradation on paths from S to A was on the interdomain link between S and A. This simplified AS-level tomography approach relies on three key assumptions for this inference to be correct:

(1) **Assumption 1: There is no congestion internal to ASes.** Thus, inferred congestion on end-to-end paths is at an interdomain link between ASes. This assumption is crucial to tomography at the AS-level; because finer-grained path information is not available (or not used), the internal structure of ASes is unknown. This assumption relies on the internal networks of ASes being well-provisioned to handle all incoming or outgoing traffic.

(2) **Assumption 2: The server and client ASes directly interconnect.** That is, no other ASes are on paths from the server AS to the client AS. If both Assumption 1 and 2 hold, the tomography problem has a straightforward solution: any observed congestion on paths from server AS S to access AS A must be on the AS-level interconnection between S and A.

(3) **Assumption 3: All router-level interconnections over which an inference is made for the AS interconnection behave similarly.** If this assumption holds, then AS-level inference accurately reflects the performance of every link

or router-level interconnection within it. For this assumption to hold, it is critical that throughput measurements from a server in AS S to clients in AS A are aggregated in a way that the paths to those clients traverse a single interconnection between S and A, or interconnections between S and A that behave similarly (e.g., parallel links between the same border routers). As Claffy et al. [14] discuss, interdomain congestion often shows regional effects. Consequently, aggregation across interdomain links is particularly problematic if those links are in different geographical regions, as they could vary widely in terms of diurnal throughput patterns.

These assumptions bear careful consideration in today's complex Internet ecosystem. The first assumption, that ASes do not experience internal congestion, derives from historical knowledge that networks generally are willing to invest significant capital in upgrading their own internal infrastructure, while disputes tend to arise at interconnection points where who pays for the upgrade depends on private negotiation between the interconnecting parties. The data at our disposal does not allow us to investigate this assumption; however evidence from recent events suggests that it may be valid. In Section 4 we use the public M-Lab path data (though limited) to test the validity of assumptions 2 and 3.

4 IS THE TOPOLOGY AMENABLE TO SIMPLIFIED TOMOGRAPHY?

Assessing the validity of the two topological assumptions described in Section 3.1 (2 and 3) will inform our judgment of the feasibility of applying simplified AS-level tomography without the use of finer-grained path information. A limited set of path measurements in the M-Lab dataset sheds doubt on whether these assumptions hold generally. A fundamental requirement in our analysis is to use traceroute measurements from M-Lab to reason about AS-level interconnections between networks. Luckie, et al. [25][2] describe in detail many things that can go wrong in inference of boundaries between ASes. Two recent pieces of work, bdrmap [26] and MAP-IT [28] have taken some steps toward overcoming those challenges. In this section, we use MAP-IT, the more general of the two interdomain link identification tools which can use a set of traceroutes that has already been collected, and apply it to the set of traceroutes collected from the M-Lab platform. We first describe the constrained set of path information available from the M-Lab, and use MAP-IT to analyze whether this limited data reveals the extent to which the M-Lab server and client are in adjacent ASes; We then apply MAP-IT to the collected traceroute data to analyze whether NDT measurements from an M-Lab server often reach a given access AS over the same physical interconnection link. For most of this analysis we use M-Lab data from 2015 in order to align it with the time periods in which the M-Lab reports and "Battle for the Net" [42] studies were released (Nov 2014 to May 2015).

4.1 Path measurements in the M-Lab dataset

Each site in M-Lab's platform is configured to launch a Paris traceroute toward every client that initiates any TCP-based measurement to the M-Lab test server. For each NDT measurement a client initiates, the server should launch a Paris traceroute toward the client. In M-Lab's initial Paris traceroute deployment, the infrastructure ran this traceroute service as a single-threaded process; consequently, if the server was performing a traceroute to client c_1 while client c_2 generated a new NDT measurement, the server would not perform a traceroute toward client c_2.[3] The platform also does not explicitly associate an NDT measurement with its corresponding Paris traceroute; the only way to match NDT measurements to corresponding traceroutes is to search the data for Paris traceroutes that were executed closely after a client ran an NDT measurement. To perform this association in the data, we matched NDT tests run by each client with the first traceroute from the server to that same client within a 10-minute window after the NDT test. With this window, the available traceroutes during May 2015 allowed us to match 71% (527,480 out of 743,780 NDT tests) from clients to M-Lab servers (with both endpoints in the U.S.). If we relaxed the matching window to allow traceroutes either before or after the NDT test, we were able to match 87% of NDT tests with traceroutes. Given that the incomplete matching was a known issue for the M-Lab team, we analyzed the NDT and Paris traceroute data from March 2017 to check whether the matching fraction had improved. We found that in March 2017, we were able to match about the same fraction, 76% of NDT tests (4,689,239 out of 6,185,394) from U.S. M-Lab servers to U.S. clients using a window of 10 minutes after the NDT test.

Another limitation of M-Lab's traceroute support is that that traceroutes are only in one direction (server to client). Clients usually run the NDT client using the web-browser implementation, where the client cannot traceroute to the server because traceroute requires lower level access to sockets. Consequently, paths from clients to M-Lab servers are not visible in this data set.

4.2 Investigating Assumption # 2: Are servers and clients in adjacent ASes?

Using the corpus of traceroute data from M-Lab in May 2015 that we matched with NDT tests (Section 4.1), we investigated Assumption 2, i.e., that server and client ASes were generally directly connected. We extracted traces from all U.S. M-Lab servers to clients in 12 major U.S. ISPs listed in the Measuring Broadband America report [20]. To identify clients in various ISPs, we used the prefix-to-AS mapping from CAIDA's AS-rank project [12], which used public BGP data from 1-5 May 2015.

Identifying interdomain links in traceroute:
In order to identify whether the server AS and client AS are directly connected or not, it is necessary to identify the AS boundaries in the traceroutes. Identifying the interdomain link between the server AS and client AS in a traceroute path faces several challenges [25]. One is that in a transition between ASes A and B, the interdomain link interfaces could be numbered out of either A or B's address space.

[2] Section 3.2 of that paper: "Inferring Interdomain Links".

[3] M-Lab corrected this issue in 2016 [24], which however led to further issues described in Section 2.2.

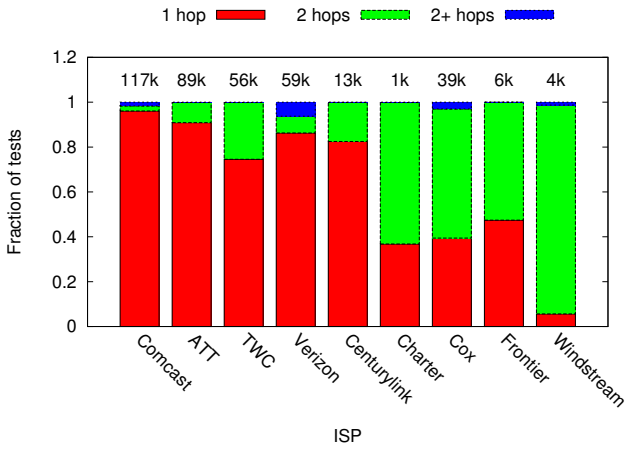


Figure 1: AS hops traversed in traceroute paths to clients in 9 large access ISPs running measurements on M-Lab in May 2015. The number above each bar denotes the number of traceroutes matched with NDT tests from M-Lab servers toward that ISP. The largest U.S. ISPs are mostly directly connected to the ASes hosting M-Lab servers. Charter, Cox, and Frontier are notable for having a smaller fraction of measurements that traverse just one AS hop. Assumption 2 does not always hold for these ISPs.

There are further challenges such as third party addresses that appear in traceroute that may confuse the identification of AS boundaries. Finally, we do not control the source or the destination of the traceroute, and therefore cannot perform additional measurements.

Fortunately, a recent effort by Marder et al. [28] focused on exactly the problem of inferring interdomain links in a set of traceroutes that have already been collected. The MAP-IT algorithm works on the basic premise that a single traceroute is insufficient to identify the AS borders that were traversed. Instead, collating together multiple traceroutes provides more constraints that can be used to infer which interfaces represent an interdomain link. The MAP-IT algorithm uses traceroutes along with additional information such as prefix-AS mappings, AS relationship data, AS-to-Organization data, and list of IXP prefixes to infer each AS boundary. This approach effectively handles the challenges posed by naming point-to-point interfaces from /30 or /31 prefixes, minimizes the impact of third-party addresses and load balancing, and corrects for mistaken inferences due to low visibility of certain interfaces in traceroute paths. Marder et al. showed that the algorithm achieved more than 90% accuracy on the datasets they tested.

We processed the entire set of matched traceroutes from May 2015 through the MAP-IT [28] algorithm. In addition to the traceroute-derived adjacencies, we used CAIDA's prefix-AS mapping derived from BGP routing tables from May 1-5, 2015, CAIDA's AS-Organization mapping [13] from July 2015 (the closest available snapshot to May 2015), and a list of IXP prefixes obtained from peeringDB [34] and PCH [32] as input to MAP-IT. For each Paris traceroute from the server to client, we then used the inference from MAP-IT to determine if the server AS and client AS were directly connected, or whether there were additional interdomain links between the server and the client AS. We considered sibling ASes as the same AS hop using information from CAIDA's AS-to-Organization dataset [13]. To obtain sibling AS lists for the client ASes, we used a manually curated list of sibling ASes for each of the top U.S. ISPs that we considered. To curate this list, we used CAIDA's AS-to-Organization dataset, Hurricane Electric's set of BGP tools [22] and then manually inspected the resulting set to remove false positives (ASes that were not siblings).

ISP	Number of subscribers (Q3 2015)
Comcast	23,329,000
AT&T	15,778,000
Time Warner Cable	13,313,000
Verizon	9,228,000
CenturyLink	6,048,000
Charter	5,572,000
Cox	4,300,000
Cablevision	2,809,000
Frontier	2,444,000
Suddenlink	1,467,000
Windstream	1,095,100
Mediacom	1,085,000

Table 1: Broadband access providers in the United States with more than one million subscribers as of Q3 2015 (retrieved from Wikipedia [2] page history)

Analyzing connectivity between Server and Client ASes: We found that 82% of the 383k traces we could analyze toward the 12 ISPs had the server AS connected directly to the client AS. However, this fraction varied considerably by ISP: 91% for AT&T, 96% for Comcast, 82% for CenturyLink, 86% for Verizon, 75% for Time Warner Cable, but only 37% for Charter, 39% for Cox, and 47% for Frontier (Figure 1). In particular, Charter, Cox, and Frontier were all in the top 10 ISPs in the U.S. in Q3 2015, yet had a much smaller fraction of tests that traversed a single AS hop from the server to client. Table 1 lists the broadband access providers in the US with more than one million subscribers as of Q3 2015. Correlating these numbers with Figure 1, we find that the top 5 broadband providers — Comcast, AT&T, Time Warner Cable, Verizon and CenturyLink — had a high fraction (greater than 80% for all except Time Warner, and greater than 90% for Comcast and AT&T) of observed paths with just one AS hop from the server to client. The fractions were lower for ISPs ranked between 5 and 10 (Charter, Cox, and Frontier). Windstream was ranked 11 in terms of subscribers in Q3 2015, and had only 6% of tests that traversed a single AS hop.

It is important to note that M-Lab servers are hosted in commercial networks; the connectivity between those networks and broadband access providers is driven by the economic incentives of those ASes, and all networks hosting M-Lab servers may not choose to connect directly with all access providers, as we observed in our data. Indeed, we find that even for the top 5 ISPs, there is a small fraction of tests that traverse one (or even two) AS hops between server and client. These cases are due to M-Lab servers in networks that do not have direct peering agreements with those access ISPs.

Summary: We conclude that the assumption of direct connection between server AS and client AS during May 2015 appeared to be true for the top 5 U.S. residential broadband access providers as of 2015, and not always true for 3 of the top 10 providers. Clearly, when analyzing NDT tests between a given server and client AS, care must be taken to ensure that the server and client AS are directly connected, using traceroutes and a technique to identify AS boundaries in traceroutes. Given the dynamic nature of AS-level interconnection, these conditions merit periodic re-examination.

4.3 Investigating Assumption # 3: diversity of interconnection to access providers

As discussed in Section 3.1, the simplified AS-level tomographic approach used in the original M-Lab report [27] implicitly assumes that either a) all measurements between that server to clients in the access AS traverse a single IP link or router-level interdomain interconnection, or b) that all IP links or router-level interconnections traversed by those measurements are similar in performance. These assumptions are required because ideally tests should not be aggregated across multiple links; if they are aggregated, they should be across links that are likely to behave similarly. Claffy et al. [14] discuss that interdomain congestion often shows regional effects. Consequently, aggregating tests across links is particularly problematic if those links are in different geographical regions, as they could vary widely in terms of diurnal throughput patterns. The M-Lab service uses proximity-based server selection to try to ensure a client performs its measurement to the geographically closest M-Lab server. We investigate the validity of Assumption 3 by exploring the topological diversity of interconnection links between an M-Lab server and NDT clients in an access ISP AS, i.e., the set of IP-level interdomain links traversed in tests from a server to client.

Identification of IP-level interdomain links

In Section 4.2, we used MAP-IT to identify the interdomain links in traceroute paths from May 2015 for the purposes of AS adjacency analysis. We reuse that same dataset to investigate the diversity of router-level interconnection, as it contains all the information necessary to identify the IP-level interdomain link traversed in a traceroute path. Specifically, for NDT tests (which could be matched with a corresponding Paris traceroute) from a server in AS S to clients in access AS A, we examine the traceroutes and determine (using MAP-IT) which IP-level interdomain links those traceroutes traversed.

Fine-grained link-level topological analysis

Our results confirmed that AS-level aggregation of measurements masked the diversity of interconnection between ASes. Table 2 lists the number of interdomain links observed from an M-Lab server in Atlanta hosted by Level 3 to 6 access ISPs, and the number of NDT measurements performed across all observed interconnections with that access ISP in May 2015. The third column lists the number of NDT tests that traversed each interdomain IP link between Level3 and that ISP. Only a single ISP, Frontier, has a significant number of tests (107) that cross a single interdomain IP link. All paths to other ISPs either have a small representation of measurements (< 100), or cross multiple interdomain IP links. Distribution of measurements across interdomain links is not uniform. Comcast's AS22909 had 1,130 measurements traversing one interdomain IP link, while Comcast's AS7922 had a total of 1767 measurements traversing two interdomain IP links (with an uneven distribution). Overall, we found that the tests to Comcast traversed 18 different AS-level links between Level3 and Comcast, comprising 30 unique IP links. We found that a majority of measurements (2395) to AT&T (AS7018) traversed a single IP link in Atlanta (we found the geographic location using reverse DNS lookups of the inferred interdomain hop), the next highest number (820) crossed an IP link in Washington DC, and 770 measurements crossed an IP link in New York City. Therefore the assumption that measurements aggregated at the AS level reflect a single connection between the server and client ASes in a given geographic region does not hold in all cases. The observed geographical spread of the interdomain links is especially problematic given the possibility of regional congestion effects [14].

Client ISP (ASN)	# Links	# NDT tests per link
Comcast (AS7922)	2	1759,8
Comcast (AS7725)	1	1650
Comcast (AS22909)	1	1130
AT&T (AS7018)	14	2395,820,770,216,137, 25,21,19,19,17,17,8,2,1
Verizon (AS701)	8	548,62,54,42,20,2,1,1
Verizon (AS6167)	2	3,3
Cox (AS22773)	39	total 817, max 378
Frontier (AS5650)	1	107
CenturyLink (AS209)	4	383,39,22,1

Table 2: Interdomain links to top U.S. ISPs seen by M-Lab server atl01 (Level 3) in Atlanta (May 2015), with the number of tests traversing each link. We only show the top 3 ASN borders from Level 3 to Comcast with the highest number of tests – in reality the data showed 18 unique AS-level links between Level 3 and Comcast, and 30 unique IP-level interdomain links distributed across these ASNs. Distribution of tests across interdomain links is not uniform: Comcast's AS22909 had 1,130 tests traversing one interdomain IP link, while Comcast's AS7922 had 1767 tests traversing two interdomain IP links.

A limitation of the MAP-IT algorithm is that it does not operate at the router level, and hence cannot reveal the presence of parallel IP links between the same pair of border routers. We used DNS names to resolve interdomain IP links into router-level interconnects for the 39 inferred interdomain links from Level 3 to Cox (AS22773), which seemed an abnormally high number. Of those 39, 12 interdomain interfaces in the level3.net domain had DNS names "COX-COMMUNI.edge5.Dallas3.Level3.net" that hinted that they were parallel links to Cox from the same Level3 router in Dallas. Another 5 IP links with the same DNS name "COX-COMMUNI.ear1.SanJose3.Level3.net" indicate that these were on a single router in San Jose. DNS entries indicated that there were two more groups of parallel links in Washington D.C. (7 links) and Los Angeles (9 links).

Summary

Based on our analysis, we conclude that aggregating NDT throughput measurement results at an AS granularity masks the fact that

different measurements could cross different IP-level links, sometimes in different geographical regions and which may have vastly different performance characteristics. Routing policies and varying client vantage point diversity can lead to significant differences in the number of measurements traversing different interconnects. On the other hand, aggregating measurements that cross different IP links between the same pair of routers may be acceptable, as load balancing generally ensures an even distribution of flows across parallel links. The range of possible scenarios highlights the importance of inferring the set of IP or router-level links that comprise the AS-level aggregation. Once the set of all IP links traversed by measurements from a server AS to a client AS are identified, it is possible to separate the NDT tests according to the IP link traversed, and evaluate whether different IP links comprising an AS-level aggregate do indeed show similar behavior. Unfortunately, the complexity of router-level interconnection may render path information from Paris traceroute insufficient to accurately identify the interdomain connection between two networks (the MAP-IT algorithm could fail or produce an incorrect inference). We need dedicated tools such as bdrmap [26] running on the server-side infrastructure to map interdomain borders, which could utilize additional measurements beyond traceroutes (e.g., alias resolution), and traceroutes in both directions associated with an NDT test, to accurately pinpoint the interdomain link traversed by each NDT test.

5 PLACEMENT OF TESTING SERVERS

Placement of servers for throughput testing has the primary objective of minimizing latency to the client (§ 2). We propose two additional considerations for using these measurement infrastructures to infer congestion on interdomain links. First, paths from within the access ISP to the test servers should cover as many interconnections of the access AS as possible. Second, measured paths should be representative of paths that *normal, user-generated* traffic from the clients traverse. We estimate, for two throughput-measurement platforms – M-Lab and Ookla's Speedtest.net – the set of interdomain interconnections of an access network that are *covered*, *i.e.*, whether a test to *any* server from that platform run from a client in the access network would traverse a given interdomain link of that access network.

5.1 Methodology to assess coverage

Measuring interdomain connectivity of access ISPs:
To measure the coverage of interdomain interconnections of access ISPs that the currently deployed server-side measurement infrastructure can provide, we first need to identify the set of interdomain interconnections of those access ISPs. For this purpose we take a different approach from that in Section 4, where we had no option but to use existing traceroutes from M-Lab servers to clients in access ISPs. Here, we use vantage points inside access ISPs to launch comprehensive topology measurements *outward* toward the whole Internet. CAIDA operates a large measurement infrastructure consisting of more than a hundred Archipelago (Ark) [11] vantage points, many of which are hosted by access networks of interest. For this study, we employed 16 Ark vantage points (VPs) located in 9 access ISPs in the U.S.: 5 in Comcast, 3 in Time Warner Cable, 2 in Cox, and one each in Verizon, CenturyLink, Sonic, RCN, Frontier, and AT&T. These vantage points are located in 8 of the top 10 broadband access providers in the U.S.; we have at least one VP in each of the top 5 providers. We focused on VPs in the U.S. for two reasons. First, M-Lab's focus is predominantly U.S.-centric. Second, recent disputes about congestion at interdomain links of access ISPs focused on U.S.-based access networks, and reports released by M-Lab [27] focused on U.S.-based networks.

To compile the set of interdomain interconnections of a given access network visible from an Ark VP in that network, we utilized *bdrmap* [26], an algorithm that accurately (the authors of [26] validated the algorithm to more than 90% accuracy on their ground truth data) infers all interdomain interconnections of a VP network visible from that VP. In the collection phase, *bdrmap* issues traceroutes from the VP toward every routed BGP prefix, and performs alias resolution (from the VP itself) on IP addresses seen from that VP in the traceroutes. We performed the data collection for *bdrmap* from our set of VPs in January and February 2017. In the analysis phase, we ran *bdrmap* using the collected topology data along with AS-relationship inferences from CAIDA's AS-rank algorithm for January 2017 [12], and a list of address blocks belonging to IXPs obtained from PeeringDB [34] and PCH [32]. *bdrmap* outputs a set of interdomain interconnections for each VP, *i.e.*, a set of border routers and neighboring networks, annotated with the type of routing relationship (customer, provider, peer, or unknown) between the VP network and the neighbor.

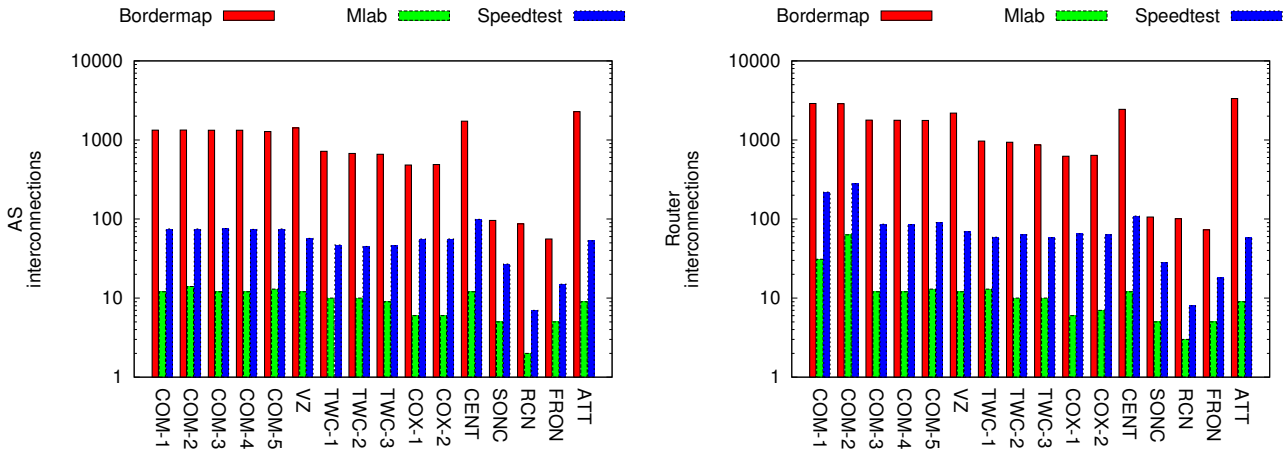
Table 3 shows, for each Ark monitor from which we ran *bdrmap*, the number of interdomain interconnections discovered at the AS and router level. We also classify the AS interconnections as customer, provider or peer using the aforementioned AS-relationship data. The data reveals the interconnection diversity in this set of access providers; some access providers such as AT&T, Verizon, Comcast and CenturyLink also operate large transit networks with thousands of customers and tens of peers. More importantly, the data highlights the scale of interdomain interconnection between large access networks. The largest access networks have hundreds of interdomain interconnections at the router-level. Even a relatively small provider such as RCN has 87 interconnections at the AS-level and 101 at the router-level.

Measuring the coverage of interdomain links
To ascertain the set of interdomain links that were *covered* using the M-Lab or Speedtest.net servers, we performed traceroutes from each Ark VP toward each of the M-Lab and Speedtest.net servers. We use the output of *bdrmap* to identify the interdomain link, if any (at both the router and AS-level) traversed by the traceroute. If the traceroute from a VP to a testing server S traverses a router-level interdomain link r corresponding to the AS-level link A, then we classify AS A and the router-level interconnection r with AS A as *covered* by the server S.

Measuring the paths to popular web content
We also wanted to ascertain the intersection between the interconnections that are covered using either the M-Lab or Speedtest.net server infrastructure, and those on the paths toward popular web content from each access ISP. For each domain in the Alexa top 500 U.S. sites [3], we scraped the default page and extracted all subdomains. We performed DNS lookups of those domains at the

Network	Ark VP	ALL borders		CUST borders		PROV borders		PEER borders	
		AS	Router	AS	Router	AS	Router	AS	Router
Comcast	bed-us	1333	2896	1115	1738	3	37	41	541
	mry-us	1336	2874	1118	1740	3	43	41	478
	atl2-us	1327	1785	1107	1318	3	20	41	139
	wbu2-us	1050	1485	897	1129	4	23	48	131
	bos5-us	1279	1768	1070	1293	3	16	40	159
Verizon	mnz-us	1423	2187	1304	1988	12	32	21	49
TWC	ith-us	720	968	588	662	3	28	28	83
	lex-us	676	935	547	613	3	29	27	83
	san4-us	660	865	535	599	3	26	28	65
Cox	msy-us	482	623	363	410	4	13	21	27
	san2-us	488	639	370	424	4	15	21	29
CenturyLink	aza-us	1729	2439	1572	2186	3	7	42	99
Sonic	wvi-us	96	106	6	6	4	5	10	10
RCN	bed3-us	87	101	35	38	1	5	36	41
Frontier	igx-us	56	73	29	30	3	6	17	29
AT&T	san6-us	2283	3336	2123	2872	12	127	40	132

Table 3: Statistics from our border identification process. We ran *bdrmap* in Jan-Feb 2017 on a wide variety of networks in terms of size. While each of the measured networks provides broadband access, several networks such as AT&T, Verizon, CenturyLink and Comcast provide transit, which is reflected in the large number of AS customers. From the point of view of congestion measurement, the number of peers (and particularly the number of router-level peer interconnections) is important.



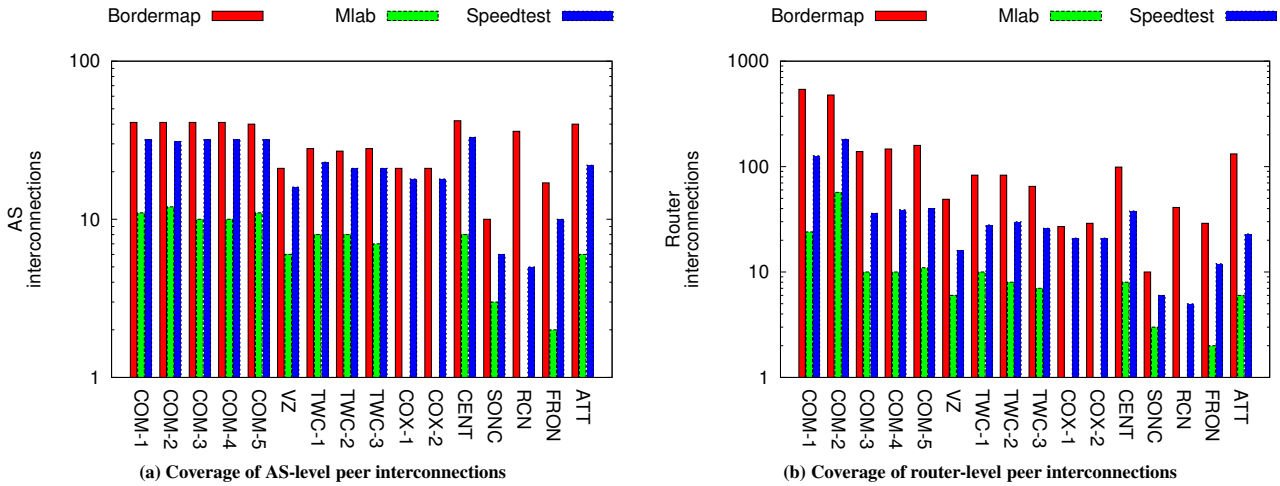
(a) Coverage of AS-level interconnections

(b) Coverage of router-level interconnections

Figure 2: Per Ark VP, AS-level (left graph) and router-level (right graph) interdomain interconnections discovered by bdrmap, and number of those interconnections appearing in traceroutes to M-Lab and Speedtest.net servers in January 2017. Across Ark VPs, only a few AS and router-level interconnections discovered by bdrmap were covered using M-Lab servers. Speedtest.net servers provided better coverage of both AS and router-level interconnections than M-Lab.

VP, to resolve the extracted domain's IP addresses. The resolved IP addresses differ per VP because we use the DNS server of the ISP hosting the VP. We refer to this set of IP addresses as the *Alexa targets*. We then performed traceroutes from each VP toward each Alexa target IP address in our list, as well as to all M-Lab and Speedtest.net servers. We processed the traceroutes toward Alexa targets, M-Lab servers and Speedtest.net servers, using the output of bdrmap to identify both router-level and AS-level interdomain interconnections of the VP network traversed on those paths.

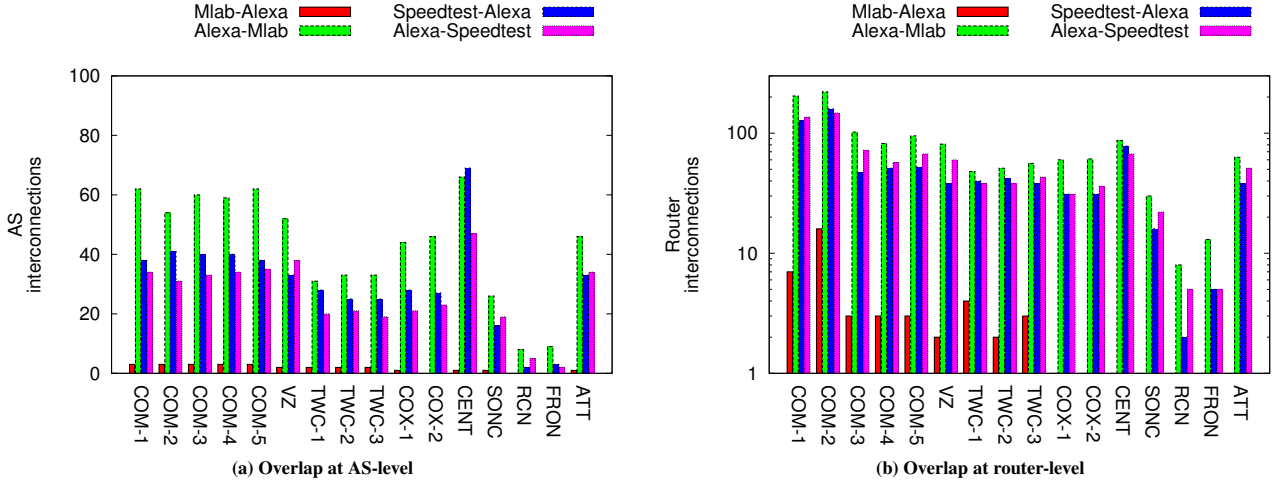
We acknowledge that a limitation of this methodology is that we use paths from within the access ISP toward the testing servers and content sources, and do not have visibility into paths in the opposite direction. Previous studies have shown, however, that path asymmetry at the AS-level is significantly less pronounced than at



(a) Coverage of AS-level peer interconnections

(b) Coverage of router-level peer interconnections

Figure 3: Per Ark VP, AS-level (left graph) and router-level (right graph) peer interconnections discovered by bdrmap, and the number of those interconnections appearing in traceroutes to M-Lab and Speedtest.net servers in January 2017. Across VPs, only a subset of peer interconnections are covered using M-Lab and Speedtest.net. Speedtest.net servers provided better coverage of both AS and router-level peer interconnections than M-Lab.



(a) Overlap at AS-level

(b) Overlap at router-level

Figure 4: Differences in the number of interconnections traversed on paths to M-Lab and Speedtest.net servers vs. those on paths toward Alexa targets. "Mlab-Alexa" denotes the number of interconnections in traceroutes to M-Lab servers but not in traceroutes to Alexa targets. "Alexa-Mlab" denotes the number of interconnections in traceroutes to Alexa targets but not in traceroutes to M-Lab servers. The remaining two bars compare the overlap between interconnections on paths to Speedtest.net servers and Alexa targets. For each VP a significant number of interconnections on paths to popular web content were not covered using M-Lab or Speedtest.net servers.

the router-level [36]. Hence for the purpose of examining the coverage of an ISPs AS-level interconnections, we believe outbound traceroutes are sufficient. In future work we plan to use the Reverse Traceroute [23] and Sibyl [16] systems when they become available to infer inbound paths to our Ark VPs. A further caveat of our methodology is that it necessarily measures popular web content, and does not include the CDN locations from which popular videos may be served. We leave an examination of paths toward the sources of popular video content to future work.

5.2 Coverage of interdomain interconnections

Figure 2 compares the set of (AS-level and router-level) interconnections of the 16 VPs observed in traceroutes toward M-Lab and Speedtest.net targets with the set of interconnections bdrmap

discovers from those VPs. In the data we analyzed from January 2017, M-Lab and Speedtest.net servers provided coverage of a small fraction of interdomain interconnections observable from the VP. Between 0.4% (for AT&T) and 9% (for Frontier) of AS-level interconnections discovered by bdrmap for different access networks were covered using M-Lab servers, while between 2.3% (for AT&T) and 28% (for Sonic) of AS-level interconnections were covered using Speedtest.net servers. In particular, the coverage of interdomain interconnections using M-Lab servers was low for the largest U.S. ISPs — 0.9% for Comcast, 0.8% for Verizon, 1.3% for Time Warner, 1.2% for Cox, 0.4% for AT&T, and 0.7% for CenturyLink. The coverage of AS-level interconnections was higher using Speedtest.net servers due to the much larger number of servers as compared to M-Lab — 5.6% for Comcast, 4% for Verizon, 6.7% for Time Warner, 11.5% for Cox, 2.3% for AT&T and 5.7% for CenturyLink.

However, the AS-level interconnections discovered by bdrmap include many customers, especially for large transit networks like Comcast, Verizon, AT&T, and CenturyLink. Figure 3 is similar to Figure 2 but reflects only interconnections inferred as peers by CAIDA's AS-rank algorithm [12]. Arguably, settlement-free (or paid) peers are more important than customers or providers from the perspective of interdomain congestion and performance; the responsibility for upgrading congested customer-provider links lies solely with the customer, while the responsibility is less clear in the case of peers. Both M-Lab and Speedtest.net provided better coverage of peer interconnections than they did of all interconnections. Further, Speedtest.net servers provided better coverage of both AS and router-level peer interconnections than M-Lab. For example, M-Lab servers were able to cover 12 of 41 of Comcast's peer ASes discovered by bdrmap; 32 of those peers were covered using Speedtest.net servers. Other networks had similar coverage: between 2.8% (RCN) and 30% (Sonic) of AS-level peer interconnections were covered by M-Lab servers, and between 14% (RCN) and 86% (Cox) using Speedtest.net servers. At the router-level, between 2.4% and 30% of peer interconnections were covered using M-Lab servers while between 12% and 78% were covered using Speedtest.net servers.

These statistics suggest that placing throughput-based test results in the right context requires knowing what fraction of interdomain interconnections of an access network a platform can measure. While M-Lab provides an invaluable server-side measurement infrastructure to support a number of measurement tests, a comprehensive view of interdomain interconnections of access networks requires substantially more server-side coverage than M-Lab provides. More generally, building a measurement infrastructure that will provide visibility into all or even most of such connections requires topology-aware deployment of measurement servers.

5.3 Overlap with interconnections used to access popular web content

Another factor to consider when designing a measurement infrastructure to capture interconnection performance is which interconnections are traversed on paths to popular web content. Figure 4 shows, per Ark VP, the overlap between the set of interdomain interconnections covered using M-Lab and Speedtest.net servers and those traversed on paths to popular web content (the Alexa targets described in Section 5.1). For 13 of 16 VPs, we observed AS-level interconnections on paths to M-Lab servers that were not on paths to any Alexa targets. For those 13 VPs, between 8% and 25% of AS-level interconnections on paths to M-lab servers were not on paths toward any Alexa targets. More importantly, for each VP we observed AS-level interconnections on paths toward Alexa targets that were not covered using M-Lab or Speedtest.net servers. Specifically, between 79% and 90% of AS-level interconnections on paths from Ark VPs to Alexa targets were not covered using M-Lab servers. For example, in the case of our Comcast VP in Bedminster MA (bed-us), 71 AS-level interconnections were traversed on paths towards Alexa targets, of which 62 (13 peers, 28 customers, 1 provider, 20 with unknown relationships) were not covered by M-Lab servers. For the same Comcast VP, 34 AS-level interconnections (3 peers, 20 customers, and 11 unknown) out of 71 AS-level interconnections on paths to Alexa targets were not covered using Speedtest.net servers. The number of AS-level interconnections on paths to Alexa targets that are *not covered* is lower for Speedtest.net than M-Lab, indicating that the larger deployment of Speedtest.net servers provides better coverage of interdomain interconnections traversed on paths to popular web content than M-Lab. However, Speedtest.net is a closed proprietary platform and unlike M-Lab, does not support custom measurement tools.

5.4 Changes over time

We conducted the entire set of previously described measurements and analysis — bdrmap to identify interdomain borders, Alexa lookups, and coverage analysis of interdomain connections using M-Lab and Speedtest.net servers — in two snapshots, October 2015 and the more recent snapshot from February 2017 described earlier in this section. Between the two snapshots, interestingly, the number of M-Lab servers was exactly the same — 261. Speedtest, on the other hand, expanded their server footprint from 3591 (October 2015) to 5209 (February 2017). However, we found that the coverage of all AS-level interconnections using both M-Lab and Speedtest servers actually decreased by a small amount ($< 5\%$) for all ISPs. We dig deeper into changes in the coverage specifically for peer connections because, as stated earlier, those are more important from the point of view of interdomain congestion. We observed the following changes in the coverage of peer AS interconnections with Speedtest between October 2015 to February 2017: from 69% to 78% for Comcast, from 81% to 76% for Verizon, from 84% to 86% for Cox, from 63% to 55% for AT&T, from 80% to 79% for CenturyLink. Apart from the increase in the coverage of Comcast and Cox's peer interconnections, the coverage of other networks decreased. For M-Lab the corresponding numbers were: 21% to 27% for Comcast, 31% to 29% for Verizon, 13% to 5% for Cox, 28% to 15% for AT&T, and 23% to 19% for CenturyLink. For M-Lab too we find that that the coverage of Comcast's peers increased; the coverage of all other networks decreased. This analysis reiterates our earlier observation that the strategic placement of testing servers is important to achieve testability of interdomain interconnections.

6 STATISTICAL CHALLENGES

End-to-end throughput-based measurement to detect congestion involves two steps: the measurement itself, and aggregating measurements to infer congestion on the path. The analysis relies on two assumptions: (1) Internet traffic has diurnal patterns, and a link is unlikely to be persistently congested all day. (2) a client is typically limited by the access link capacity, *i.e.*, links upstream of the access link are typically not the throughput bottleneck. Therefore, if client achieves significantly less throughput during peak times than during off-peak times, a plausible explanation is that the throughput is being limited by a congested link further upstream of the access link. While superficially a sound approach, the leap from observing diurnal patterns in an aggregate set of measurements to claiming congestion relies on two further—and major—assumptions: (1) the samples used across the day, and across a variety of access link configurations are comparable, and (2) there are well-understood thresholds for detecting congestion. Two factors shed some doubt on these assumptions: limitations of crowdsourced measurements, and ambiguity in what constitutes congestion.

6.1 Limitations of crowdsourcing

Crowdsourcing has advantages in terms of size and richness of resulting samples. It also has limitations:

- *Samples cannot be controlled.* Any particular home or client likely generates only one or a few samples, and their network performance may vary widely.
- *Time of day bias.* Since users manually launch tests, there are usually more runs during the day than at night, which can make the diurnal pattern difficult to discern.
- *Service plan variance.* It is difficult to get ground truth about expected performance without input from users, *e.g.*, the user's service tier, which would suggest what the user could reasonably expect from a throughput test. Even within a region, an ISP could offer service plans with capacities that vary by an order of magnitude. Such information is typically available only to access ISPs and the users themselves (although many users do not know their service tier), and web-based tests cannot automatically obtain it.
- *Home network interference.* Cross traffic on the home network, especially on Wi-Fi, could affect throughput. Previous work has shown how home wireless networks have a major impact on performance [38, 39], and how wireless performance could vary significantly even across devices within a single home.

Figure 5 reproduces a graph from M-Lab 2015 analysis [4] that shows how throughput performance varied by time of day for AT&T and Comcast users to an M-Lab server hosted in GTT in Atlanta during May 2015. The M-Lab analysis stated that *"AT&T users experienced the most consistent patterns of congestion-related degradation across measurement points on a diversity of transit ISPs, most notably on GTT for Atlanta . . . Other access ISPs such as Comcast did not display as substantial of degradation to those same sites during the same period"*.

We examine this case in more detail. Instead of tracking only median throughput as the report does, we plot the average and standard deviation of throughput, and number of samples, to illustrate the four limiting factors described above. First, we see variance is high; during off-peak for AT&T (Figure 5a), and consistently so in the case of Comcast (Figure 5b). This variance could be caused by one of these limiting factors, e.g., differences in service plan rates exacerbated by the sparseness of measurements from a single client, or even wireless issues or differences in the home network. Second, off-peak hours have significantly fewer samples—fewer than 20 in some cases—illustrating the time of day bias. Fewer samples during off-peak hours is consistent with general network usage, but makes it difficult to compare peak versus off-peak performance with statistically significant results. With so few samples, the throughput measurements could be skewed by any of the above confounding factors.

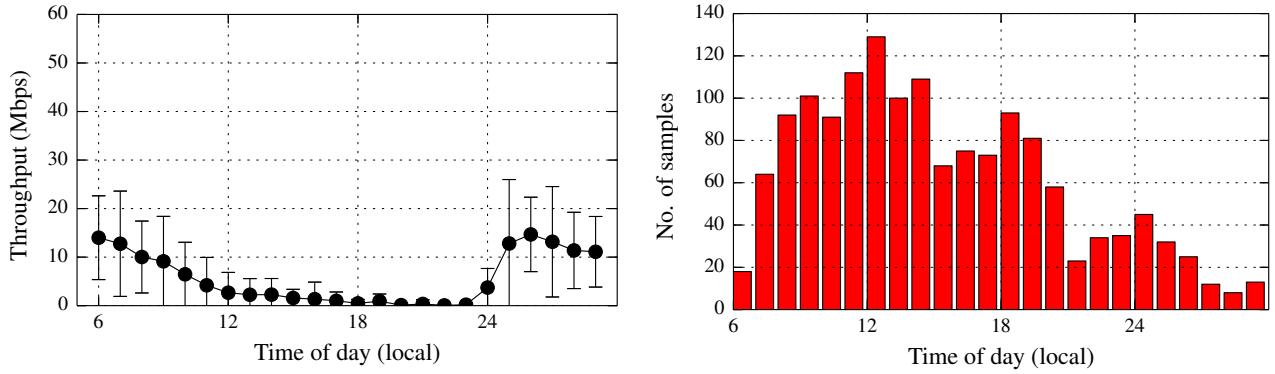
6.2 Thresholds to detect congestion

Even if we assume that we can compare peak and off-peak throughput to infer congestion during peak hours, identifying what constitutes congestion is not straightforward. The M-Lab [27] report identified examples where the peak-hour download throughput measurements dropped drastically, such as from highs of greater than 10 Mbps to less than 1 Mbps for AT&T (Figure 5a); such drops can be reasonably attributed to a link on the end-to-end path that is saturated during peak hours. However, even examples used in the report to contrast with congested links show diurnal patterns as in the case of Comcast tests to GTT servers in Atlanta (Figure 5b, which was identified as an uncongested link in the report.) In this example, the peak-to-trough difference in throughput for Comcast is about 30% (even removing the off-peak hours with few samples, this difference is 20%). Such a measurable, non-trivial diurnal throughput drop raises the question: how large a throughput drop can one safely interpret as evidence of congestion?

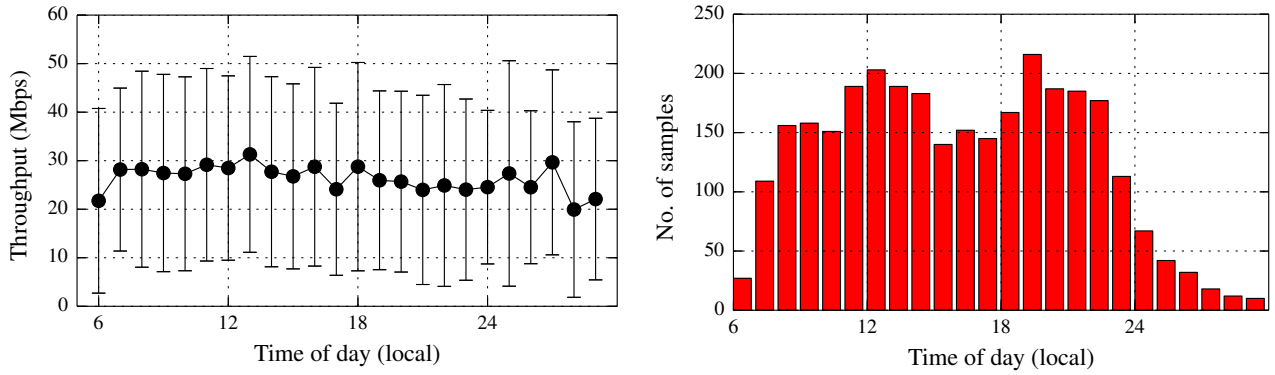
These two cases likely reflect two different link states: a link that becomes congested at peak hours vs. an uncongested link that sees higher utilization during peak hours (as most links do). For the AT&T tests in Figure 5a, the drop in throughput, coupled with very low variance, means that all tests see consistently low throughput, suggesting peak-hour congestion is the cause. It is more difficult to attribute a cause to the performance drop of Comcast tests in Figure 5b. This drop could be due to sample bias, more users sharing the cable medium during peak hours, or that a subset of Comcast users experienced lower peak-hour throughput due to contention at some point on the end-to-end path. This raises the question: is there a more direct way to identify whether a flow was congested by an already busy link or whether the flow itself drove congestion in a (presumably access) link? Distinguishing these two cases is still an open challenge in throughput-based congestion inference.

7 LESSONS LEARNED

Congestion at ISP interconnections has been a recent focus in the research, economic, and regulatory arenas. There have been recent, high-profile, efforts in attempting to understand the extent of such congestion by using crowdsourced throughput tests from distributed measurement infrastructures. We used public measurement data from these efforts, and our own measurement experiments, to investigate challenges in inferring interconnection congestion using end-to-end throughput measurements. The methodological challenges



(a) GTT server in Atlanta to AT&T customers. (May 2015)



(b) GTT server in Atlanta to Comcast customers. (May 2015)

Figure 5: Diurnal throughput (left) and number of samples (right) using NDT tests from a M-Lab server in GTT to clients in AT&T (a) and Comcast (b). AT&T users see a drop in throughput to less than 1 Mbps during peak hours. Comcast users see a drop as well, but not to the same extent. The number of samples are also much fewer during off-peak hours.

fall into three categories: the complexity and opaqueness of the Internet's topology; visibility of interconnections used to access popular content; and statistical issues associated with crowdsourced sampling of performance measurements. Overcoming each challenge with available data requires making several assumptions, and we used this broad set of measurements to assess the degree to which these assumptions hold on today's Internet.

First, pinpointing the location of congestion using end-to-end measurement requires application of network tomographic techniques to detailed router-level path information in both directions taken at the time of the end-to-end measurements. Obtaining such information is an open research and policy challenge. An alternative is to use coarser-grained, i.e., AS-level tomography, and to further simplify the tomography with three assumptions: there is no congestion internal to ASes, only at interconnects; the two endpoints of the measurement are in directly connected ASes; and there is only one physical link connecting them which the measurement traffic traverses. The first assumption is consistent with comments from many industry players we have seen in discussions on NANOG and received via personal communication; we did not have data to investigate it in this study. With respect to the second assumption, our analysis of data from M-Lab's study [27] revealed that although most clients are usually one AS hop away from M-Lab's testing server, most tests between some AS pairs traverse multiple AS hops. Having more than one AS hop between the server and client sheds doubt on congestion inferences, because any interdomain link in the path could be the point of congestion (assuming also that congestion is more likely at interdomain links than internal to networks).

The third assumption is more problematic: the limited path information available from the M-Lab study shows that the interconnections between the same two pair of ISPs are often *not* crossing the same IP link. This is consistent with recent studies that show that larger ASes tend to interconnect with each other in many locations, and congestion on these interconnections can often have regional effects [14].

Second, what we can currently measure with existing server-side measurement infrastructure is only a small subset of the interconnection landscape, and may not provide visibility of paths that carry

popular content. Our analysis revealed that the set of interdomain interconnections, both at the router and AS-level, testable using M-Lab or Speedtest.net infrastructure typically had low overlap with those traversed on the paths to popular web content.

Finally, crowdsourcing methods have advantages in the potential for sampling breadth across geographical regions, ISPs, service plans, and home network conditions. However, in practice, crowdsourced measurements can yield exactly the opposite: an uneven distribution of samples across time of day, access link speeds, and home network qualities. Another statistical challenge is selection of a threshold drop in throughput to constitute evidence of congestion.

Recommendations

Our methodology and analysis offers opportunities for measurement platforms to tune the deployment of their measurement servers to improve the coverage of relevant interdomain interconnects. We offer several suggestions to mitigate the impact of these issues and enable more rigorous inference of congestion: Most critical, every throughput-based test must include a traceroute taken as close as possible in time to the test, preferably in both directions. Deploying a router-level interconnection inference tool such as *bdrmap* [26] on a server-side infrastructure such as M-Lab would greatly increase situational awareness of the topology state during measurements by allowing inference of which router-level interconnects a given test traverses. To limit the potential of interference from multiple points of interconnection, measurement projects could strategically deploy servers to increase the fraction of one-hop tests, modifying server selection logic to select only directly connected servers, and using path information to discard tests that traverse more than one AS hop. In any case, analysis of throughput measurements should not aggregate across router-level links (particularly if the router-level links are in separate geographical regions). Doing so may aggregate across links with dissimilar performance characteristics [14]. To ensure that congestion inferences reflect performance that users actually experience, measurement platforms should incorporate regular measurements of paths to popular content. Otherwise, claims about congestion at interconnects should acknowledge that those interconnects may not be on the path from the most popular content to users. Finally, the community could mitigate the statistical limitations of crowdsourcing by using other measurement platforms to run periodic tests that complement the crowdsourced tests. Ark, BISmark, and RIPE Atlas are a few examples of platforms that support repeated longitudinal measurements. These other platforms are not provisioned to support the bandwidth requirements of NDT throughput measurements, but they, as well as M-Lab, could support lower-impact techniques such as TSLP [25] to provide additional insight into the presence and location of congestion.

Future work

We hope that these recommendations will lead to improvements in measurement platforms to better support the inference and localization of congestion. With regard to the analysis of existing data, a focus of our ongoing work is to use the insights we have gleaned from the analysis of router-level interconnection (Section 4) to more rigorously analyze the M-Lab data. In particular, we are using the NDT tests in conjunction with Paris traceroutes and MAP-IT inferences to identify the specific IP-level interconnection traversed by each test. By doing so, we will be able to analyze the performance of tests traversing each individual IP-level interconnect between a given source and client AS, and to make inferences about whether specific IP-level interconnection links are congested.

In general, being able to detect the presence and type of congestion is an still open problem. It would be useful if speed tests such as those conducted by M-Lab, various speed test sites, and the FCC Measuring Broadband America infrastructure could reveal more information about the path than simply achievable throughput. We have taken some steps in this direction with recent work [37] that uses RTT signatures extracted from speed tests to determine whether a TCP flow was limited by an already congested link in the path, or whether it started on an initially unconstrained path, thus driving buffer behavior. While this method cannot by itself pinpoint the *location* of the congested link, we believe that it can provide additional information useful for interpreting the results of speed tests. Our focus in the near future will be on getting this capability deployed on the M-Lab and FCC Measuring Broadband America infrastructure.

ACKNOWLEDGEMENTS

We would like to thank our shepherd, David Choffnes, and the anonymous reviewers for their valuable feedback. This work was supported by NSF CNS-1414177 and a grant from Google, but this paper represents only the position of the authors.

REFERENCES

[1] Internet Health Test. http://internethealthtest.org, 2015.
[2] Internet in the United States, 2017. https://en.wikipedia.org/wiki/Internet_in_the_United_States.
[3] Alexa. Top Sites in United States. http://www.alexa.com/topsites/countries/US, 2017.
[4] C. Anderson. New Opportunities for Test Deployment and Continued Analysis of Interconnection Performance. http://www.measurementlab.net/blog/interconnection_and_measurement_update, 2015.
[5] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding Traceroute Anomalies with Paris Traceroute. In *Proceedings of ACM SIGCOMM Internet Measurement Conference (IMC)*, Oct. 2006.
[6] S. Bauer, D. Clark, and W. Lehr. Understanding Broadband Speed Measurements. In *Telecommunications Policy Research Conference (TPRC)*, Oct. 2010.
[7] BISMark: Broadband Internet Service Benchmark, 2017. http://projectbismark.net/.
[8] J. Brodkin. Time Warner, Net Neutrality Foes Cry Foul Over Netflix Super HD Demands. http://arstechnica.com/business/2013/01/timewarner-net-neutrality-foes-cry-foul-netflix-requirements-for-super-hd/, 2013.
[9] J. Brodkin. Why YouTube Buffers: The Secret Deals that Make-and-break Online Video. *Ars Technica*, July 2013.
[10] S. Buckley. France Telecom and Google entangled in peering fight. *Fierce Telecom*, 2013.
[11] CAIDA. Archipelago (Ark) Measurement Infrastructure. http://www.caida.org/projects/ark/, 2017.
[12] CAIDA. AS Relationships. http://www.caida.org/data/as-relationships/, 2017.
[13] CAIDA. Inferred AS to Organization Mapping Dataset. https://www.caida.org/data/as-organizations/, 2017.
[14] k. claffy, D. Clark, S. Bauer, and A. Dhamdhere. Policy Challenges in Mapping Internet Interdomain Congestion. In *Telecommunications Policy Research Conference (TPRC)*, Oct 2016.
[15] Cogent Communications Inc. Ex Parte Filing from Cogent, DISH, Free Press, Open Technology Institute, Public Knowledge related to ATT DirecTV Merger, May 2015. "http://apps.fcc.gov/ecfs/comment/view?id=60001031493", 2015.
[16] I. Cunha, P. Marchetta, M. Calder, Y.-C. Chiu, B. Schlinker, B. V. A. Machado, A. Pescapè, V. Giotsas, H. V. A. Madhyastha, and E. Katz-Bassett. Sibyl: A Practical Internet Route Oracle. In *Proceedings of the Usenix Conference on Networked Systems Design and Implementation (NSDI)*, 2016.

[17] S. Dent. Google is Testing Internet Speeds Straight from Search. https://www.engadget.com/2016/06/29/google-is-testing-internet-speeds-straight-from-search/, 2016.
[18] N. G. Duffield. Network Tomography of Binary Network Performance Characteristics. *IEEE Transactions on Information Theory*, Dec. 2006.
[19] J. Engebretson. Behind the Level 3-Comcast Peering Settlement, July 2013. http://www.telecompetitor.com/behind-the-level-3-comcast-peering-settlement/.
[20] FCC. Measuring Broadband America Report 2014. https://www.fcc.gov/reports/measuring-broadband-america-2014.
[21] Y. Huang, N. Feamster, and R. Teixeira. Practical Issues with Using Network Tomography for Fault Diagnosis. *ACM SIGCOMM Computer Communincation Review (CCR)*, 2008.
[22] Hurricane Electric. BGP Toolkit, 2017. http://bgp.he.net.
[23] E. Katz-Bassett, H. V. Madhyastha, V. K. Adhikari, C. Scott, J. Sherry, P. Van Wesep, T. Anderson, and A. Krishnamurthy. Reverse traceroute. In *Proceedings of the USENIX Conference on Networked Systems Design and Implementation (NSDI)*, 2010.
[24] D. H. Lee. Relation Between NDT and Paris Traceroute Tests on M-Lab. https://groups.google.com/a/measurementlab.net/forum/#!topic/discuss/YxI4Z_IBb9Y, 2015.
[25] M. Luckie, A. Dhamdhere, D. Clark, B. Huffaker, and kc claffy. Challenges in Inferring Internet Interdomain Congestion. In *Proceedings of ACM SIGCOMM Internet Measurement Conference (IMC)*, Nov. 2014.
[26] M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark, and k. claffy. bdrmap: Inference of Borders Between IP Networks. In *Proceedings of ACM SIGCOMM Internet Measurement Conference (IMC)*, Nov 2016.
[27] M-Lab Research Team. ISP Interconnection and its Impact on Consumer Internet Performance - A Measurement Lab Consortium Technical Report. http://www.measurementlab.net/publications, 2014.
[28] A. Marder and J. M. Smith. MAP-IT: Multipass Accurate Passive Inferences from Traceroute. In *Proceedings of ACM SIGCOMM Internet Measurement Conference (IMC)*, 2016.
[29] Measurement Labs. M-Lab Dataset. https://cloud.google.com/bigquery/docs/dataset-mlab, 2017.
[30] Measurement Labs. NDT Data Format. https://code.google.com/p/ndt/wiki/NDTDataFormat, 2017.
[31] Ookla. How Does the Test Itself Work? https://support.speedtest.net/hc/en-us/articles/203845400-How-does-the-test-itself-work-How-is-the-result-calculated-, 2012.
[32] Packet Clearing House. Full Exchange Point Dataset. https://prefix.pch.net/applications/ixpdir/menu_download.php, 2017.
[33] J. Padhye, V. Firoiu, D. F. Towsley, and J. F. Kurose. Modeling TCP Reno Performance: A Simple Model and its Empirical Validation. *IEEE/ACM Transactions on Networking*, 2000.
[34] PeeringDB, 2017. http://www.peeringdb.com.
[35] C. Ritzo. Paris Traceroute Brownout. https://www.measurementlab.net/blog/paris-traceroute-brownout/, Apr. 2017.
[36] M. A. Sánchez, J. S. Otto, Z. S. Bischof, D. R. Choffnes, F. E. Bustamante, B. Krishnamurthy, and W. Willinger. Dasu: Pushing experiments to the internet's edge. In *Proceedings of the USENIX Conference on Networked Systems Design and Implementation (NSDI)*, 2013.
[37] S. Sundaresan, A. Dhamdhere, M. Allman, and kc claffy. TCP Congestion Signatures. In *Proceedings of ACM SIGCOMM Internet Measurement Conference (IMC)*, Nov. 2017.
[38] S. Sundaresan, N. Feamster, and R. Teixeira. Measuring the Performance of User Traffic in Home Wireless Networks. In *Proceedings of the Passive and Active Measurement Conference (PAM)*, 2015.
[39] S. Sundaresan, N. Feamster, and R. Teixeira. Home Network or Access Link? Locating Last-Mile Downstream Throughput Bottlenecks. In *Proceedings of the Passive and Active Measurement Conference (PAM)*, 2016.
[40] M. Taylor. Observations of an Internet Middleman, May 2014. http://blog.level3.com/global-connectivity/observations-internet-middleman/.
[41] Various Authors. Email thread "Comments on ISP Interconnection and its Impact on Consumer Internet Performance", 2014. https://groups.google.com/a/measurementlab.net/forum/#!msg/discuss/lwVmPrbRg0w/1CTgbNcgInIJ.
[42] Various Authors. Email thread "BattlefortheNet study", 2015. https://groups.google.com/a/measurementlab.net/d/msg/discuss/RbPb18fY_VA/P7Eil4lcgLMJ.
[43] Verizon. Unbalanced Peering, and the Real Story Behind the Verizon/Cogent Dispute, June 2013. http://publicpolicy.verizon.com/blog/.
[44] D. Vorhaus. A New Way to Measure Broadband in America. http://blog.broadband.gov/?entryId=359987, Apr. 2010.