Understanding broadband speed measurements

Steve Bauer
David Clark
William Lehr
Massachusetts Institute of Technology

Table of Contents

Understanding broadband speed measurements

Steve Bauer
David Clark
William Lehr
Massachusetts Institute of Technology

Abstract

Broadband speed has emerged as the single most commonly cited metric for characterizing the quality of broadband offerings. However speed measurements for the same service can vary significantly. These differences arise from a complex set of factors including different test methodologies and test conditions. For any testing methodology, teasing apart the end-to-end tests and attributing performance bottlenecks to constituent parts is technically challenging. While the broadband access network can be the bottleneck, significant bottlenecks arise in home networks, end users' computers, and server side systems and networks. Consequently, inferences regarding how ISP delivered speeds compare with their advertised speeds need to be undertaken with careful attention to the testing methodologies employed. Many testing methodologies are inappropriate for the purposes of assessing the quality of a broadband network.

1. Executive Summary[1]

Speed is the single most important metric of interest in characterizing the "quality" of broadband service. Perceptions about broadband quality inform regulatory policy, end-user behavior (e.g., broadband subscriptions), investments in complementary assets (e.g., content and applications), as well as the marketing, traffic management, and provisioning decisions of network operators. Consequently, it is important to understand how speed is, and should be, measured.

[1] We would like to acknowledge the support of participants in the MIT MITAS (http://mitas.csail.mit.edu) and CFP (http://cfp.mit.edu) research programs. Any opinions expressed and errors are solely the responsibility of the authors.

The goals for this paper are several. First, we explain the complexity of measuring broadband speeds. There are multiple definitions of "speed" that are potentially of interest: is it a measure of potential throughput or *capacity* or is it as a measure of *average* speed as experienced by end-users? Is the focus the broadband access service or end-to-end performance? Is the goal to diagnose a potential fault or to benchmark performance? Is the interest in a single broadband connection, a geographic market, or some larger population?

Second we document key methodological differences that account for some of the variability in popular speed measurement data sources. While it is not surprising that there is significant variability in broadband speed measurements across providers, geographic markets, access technologies, and time, it is surprising how much variation can result from methodological differences in how speed is measured. Understanding these methodological differences is important to making valid inferences from the measurement data.

For example, while a principal motivation for many in looking at speed measurements is to assess whether a broadband access ISP is meeting its commitment to provide an advertised data service (e.g. "up to 20 megabits per second"), we conclude that most of the popular speed data sources fail to provide sufficiently accurate data for this purpose. In many cases, the reason a user measures a data rate below the advertised rate is due to bottlenecks on the user-side, at the destination server, or elsewhere in the network (beyond the access ISP's control). A particularly common non-ISP bottleneck is the receive window (rwnd) advertised by the user's transport protocol (TCP). In the NDT dataset we examine later in this paper, 38% of the tests never made use of all the available network capacity.

Other non-ISP bottlenecks also exist that constrain the data rate well below the rate supported by broadband access connections. Local bottlenecks often arise in home wireless networks. The maximum rate of an 802.11b WiFi router (still a very common wireless router) is 11mbps. If wireless signal quality is an issue, the 802.11b router will drop back to 5.5mbps, 2mbps, and then 1 mbps. Newer wireless routers (e.g. 802.11g/n) have higher maximum speeds (e.g. 54 mbps) but will similarly adapt the link speed to improve the signal quality. End-users also can self-congest when other applications or family members share the broadband connection. Their measured speed will be diminished as the number of competing flows increase.

Each of the different testing methodologies we examined provides insights into network performance. We concluded however that the Ookla/Speedtest approach – which typically results in higher measured data rates than the other approaches reviewed – was the best of the *currently* available data sources for assessing the speed of ISP's broadband access service. One of the key differences that accounts for this is that the Ookla/Speedtest tools utilize multiple TCP connections to collect the measurement data which is key to avoiding the receive window limitation. These tests are also much more likely to be conducted to a server that is relatively close to the client running the test.

We expect that new hardware based speed measurements, such as those being conducted now by the FCC in collaboration with Samknows (which ran a similar study for Ofcom in the UK), will produce very useful results.[2] The hardware platform eliminates many of the potential bottlenecks (e.g. noisy wireless links and receiver window bottlenecks) that can interfere with network measurements. However, such hardware based testing will not eliminate the importance of other testing methodologies. Orders of magnitude more data can be gathered by the existing non-hardware-based tests. We also expect that some of the same tests will be conducted in both the hardware and browser based environments (e.g. we expect some of the hardware testing platforms will leverage the M-labs infrastructure running the NDT test.)

Precisely because differences do matter, it is important to understand methodological details of the tests being conducted. Unfortunately uncovering the testing details is not always easy. There are marked differences in the quality of testing documentation and disclosure. Some of the organizations made adjustments to their methods and documentation in response to our queries. This highlights a kind of Heisenberg Uncertainty Principle as it applies to this sort of research: because speed data matters, purveyors of such data may adjust the methods and published information dynamically and, potentially, strategically.

We recognize that the debate over speed is, itself, strategic and engages the interests of all market participants – policymakers, users, and ISPs alike. While we do not believe there is a single method for measuring and reporting speeds that is appropriate for all contexts, we recognize the analytic and cost benefits from narrowing the range of methods and from moving toward collective agreement on best practices. Having a few generally accepted methods for measuring broadband speeds and having clear documentation to support published measurements will aid data aggregation, analysis, and interpretation.

Furthermore, although speed matters and will continue to do so in the future to all market participants interested in assessing the quality of broadband services, it is *not* all that matters. There is a risk that the focus on speed or a particular way of measuring or reporting broadband speeds might distort market behavior. We believe promoting a more nuanced understanding of how and why speed matters offers the best defense against such adverse distortions (e.g., what applications are most likely to benefit from faster broadband services? What additional mechanisms may be available to isolate speed bottlenecks at different points along the end-to-end path? What other technical attributes should be regularly measured and reported to better characterize broadband service quality?).

In writing this paper, our intent is to contribute to this public debate and to render more accessible some of the speed measurement issues to non-Internet engineering experts. Our hope is that progress may be made via a market-mediated process that engages users, academics, the technical standards community, ISPs, and policymakers in an open debate; one that will not require strong regulatory mandates. Market efficiency and competition will be best served if

[2] See http://www.samknows.com/broadband/index.php (accessed 21 June 2010).

there is more and better understood data available on broadband speeds and other performance metrics of merit (e.g., pricing, availability, and other technical characteristics).

2. Introduction[3]

Broadband speed has emerged as the single most commonly cited metric for characterizing the quality of broadband offerings. There are now a number of sites and organizations that measure the speed (and other characteristics) of a user's broadband service (e.g., Speedtest.net, Akamai, ComScore, M-Labs, Google's Youtube service, or the FCC's own broadband.gov speed tests). The data generated by such tests is often aggregated and reported in the trade press and in government reports and plays a role both in policy formulation and in decision-making by individual consumers.[4] Consequently, how speeds are measured and how the data is interpreted can have important implications for market and policy outcomes.

In this paper, we explain why the resulting speed measurements at times vary significantly. For instance, Google's Youtube service reports an average download speed of 3.83 mbps for the United States while the Speedtest.net reports a 7.71 mbps average.[5] The differences can be even more pronounced at smaller geographic granularities. For Comcast in the Boston area, Speedtest.net reports average download speeds of 15.03 mbps while Youtube reports a 5.87 mbps average.[6] Differences in the methodologies account for most of the discrepancy. The proper interpretation is not that one test is "right" or "wrong" but rather the different tests provide different insights into the end-to-end performance under different workloads.

For any testing methodology, teasing apart the end-to-end tests and attributing performance bottlenecks to constituent parts is technically challenging. While the broadband access network can be the bottleneck, significant bottlenecks arise in home networks, end users' computers, and server side systems and networks. The dynamics and settings of TCP (the dominant transport protocol on the Internet) also play a significant role in determining the resulting speed that is measured. There is also an important question about systematic biases in user initiated speed tests. Potentially users are running those tests in a diagnostic fashion when they are experiencing

---

[3] This paper is intended to be accessible to non-experts in Internet technology, but space constraints presume a level of technical sophistication. To help bridge the gap, this paper is accompanied by web-based resources for acronyms and further details. See http://mitas.csail.mit.edu/wiki

[4] The recent U.S. National Broadband Plan (http://download.broadband.gov/plan/national-broadband-plan.pdf) for instance notes both current results and recommends expanding the set of broadband performance measurements.

[5] The exact numbers from each of these services varies over time. These results were downloaded on March 16, 2010 from http://speedtest.net/global.php and http://youtube.com/my_speed respectively.

[6] Results download on March 16, 2010 by computers in the Boston area from http://speedtest.net/global.php#0,1,1,26 and http://youtube.com/my_speed. It is unclear how closely geographic areas align for different services reporting results from smaller geographic regions like states and cities.

problems.[7] The main point is that inferences regarding how ISP delivered speeds compares with their advertised speeds need to be undertaken with careful attention to the testing methodologies employed. Many testing methodologies are inappropriate for the purposes of assessing the quality of ISP access networks.

Measuring and understanding the performance of the broadband access networks, though, remains an important policy question. We examined in detail a number of speed tests, using packet traces to understand what factors shaped the outcome and how the test was performed. We discuss a number of these cases, and argue that the Ookla/Speedtest test methodology[8] is more likely than the other tests we examine to correspond to the speed of an access link for common network usage patterns. We explain why the Ookla/Speedtest methodology often produces higher estimates of speed then other tests. We discuss limitations and some of the technical rationales underlying the tradeoffs that must be made in this type of testing.

We conclude with policy recommendations that emphasize that while speed remains an important statistic for evaluating the quality of broadband and will likely remain so, appropriate metrics for evaluating the performance of broadband services should consider additional characteristics as well. Moreover, we expect that the relevance of raw speed measurements – separate from more nuanced context-dependent considerations – will become increasingly less relevant as the speed of the typical broadband offering becomes ever faster. We also call for better documentation of speed testing methodologies from all organizations so that third-party validation of measurements is easier.

3. Defining what is meant by broadband speed

We first have to understand exactly what is meant by the "speed of broadband". The casual interpretation of speed is that it indicates "how fast you can go". More speed, everything else being equal (particularly price), is always better. But beyond this casual interpretation of speed, lie a variety of divergent meanings. For example, there is the speed implied by the way the broadband access service is configured (its "capacity"), the speed that is marketed or understood by consumers ("advertised"), or the speed that is actually experienced by users ("achieved"). We put these terms in quotes because we are using them as imprecise shorthand.

3.1. Provider configured broadband speeds

In the advertisements of broadband providers, upload and download speeds provide information about how the link between the customer's location and the broadband provider is configured. It

---

[7] For some networks such as broadband satellite networks or wireless broadband networks which often have a smaller number of test samples in a given data set, we suspect that tests run in an attempt to diagnosis a problem could account for a significant percentage of tests. However, we are personally familiar with many expert and non-expert users that simply like to measure their speed for fun.

[8] http://wiki.ookla.com/test_flow

generally corresponds to configuration settings that are set in network equipment such as cable and DSL modems and the routers or other network equipment, and are generally intended to give an indication of the maximum or *peak* data rates that a customer may experience. More typically, however, these advertised data rates are taken to imply something meaningful about the experience of using the broadband service. The challenge therefore is how to define and measure "speeds" that are both representative of the user experience on the broadband network and are robust to the messy measurement conditions of the real world.

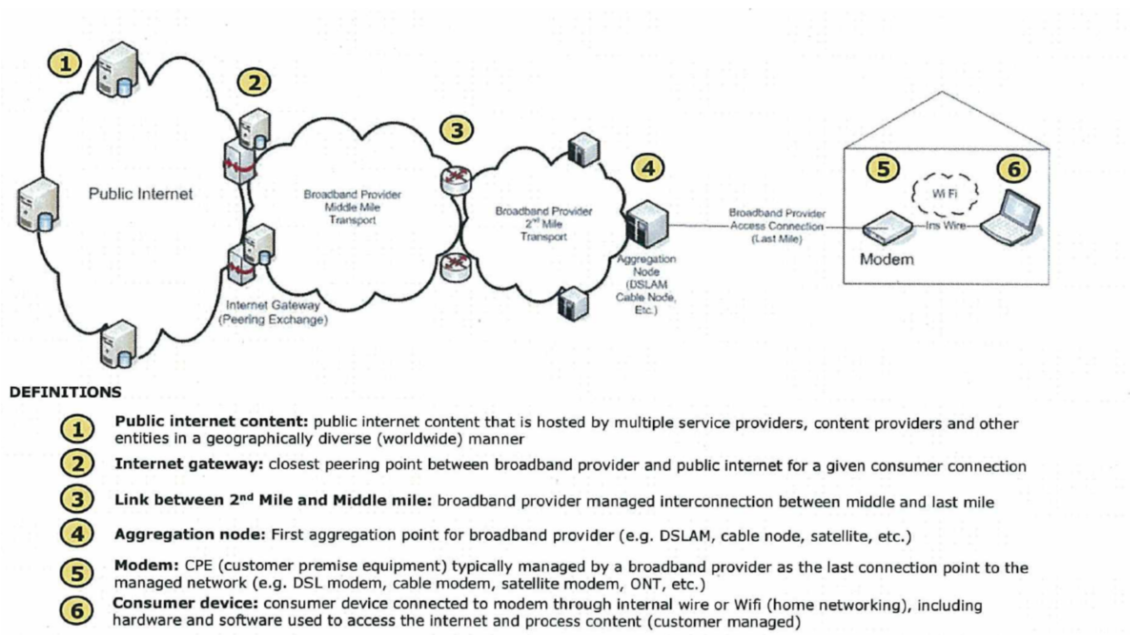

Figure 1: Network diagram showing the most important devices and links connecting broadband users with destinations on the Internet. Broadband providers service demarcation points are between points 2 and 5 generally. Most of the speed tests we examine in this paper measure the speed between 1 and 6. Diagram source: "Federal Communications Commission Request for Quotation for Residential Fixed Broadband Services Testing and Measurement Solution."

The speed advertised and set by a broadband provider (i.e. the speed between 4 and 5 in Figure 1 above) is an important metric for characterizing the broadband service and does provide a useful first-level indicator of the expected "speed" of traffic carried by the network. However, it may not be directly indicative of what a user will actually experience. There are numerous reasons why differences arise.

The most commonly assumed reason for measuring speeds lower than the configured speed is that broadband access networks rely on multiple users sharing the available network capacity (particularly the link between 4 and 5 in Figure 1 in the case of cable and between 3 and 4 in the case of DSL). Because end-user demands are unlikely to be perfectly correlated in time, the traffic of multiple users can share network capacity, thereby substantially reducing the total costs of providing network services. Assuming no problems elsewhere, a user whose connection is configured to run at a peak rate of 15Mbps can realize that speed across parts of the shared infrastructure if enough other users are not simultaneously utilizing the network.[9] The throughput during the peak usage hours depends upon how many users are configured to share (a design decision) and how many of them are simultaneously active (user behavior). This sharing is fundamental not just for broadband networks, but the Internet as a whole.[10] The potential aggregate demand from senders and receivers on the Internet vastly exceeds the capacity of the network. Sharing saves money and allows everyone to get more than they would on their own for a given price. As we demonstrate in this paper, however, congestion arising from infrastructure sharing with multiple users is *not* the most common reason that speeds lower than the provider configured speed are measured. This may change over time as other sources of performance impairments are fixed.

The configured "up to" speed advertised by a provider is also not actually the maximum speed that one can measure on a network. Higher-than-expected measurements can occur for a number of reasons. The "line speed" for a broadband service can be faster than the configured speed.[11] The line speed is the speed at which an individual packet is sent on the wire or fiber and is a property of the physical and link layer technologies employed. A packet in a DOCSIS 2.0 network will be sent at something around 38 mbps for instance.[12] Short bursts of packets can also be sent closer together than would be implied by the configured speed.[13] Consequently, the speed that may be measured over very short time intervals is complicated by non-trivial scheduling algorithms employed by the lower link-layer protocols.

However, even over longer intervals, the speed measured may be *more* than the data rate that is advertised. Access networks can remove the configured speed limits for periods of time thus allowing the users to send at higher rates. Some cable broadband networks employ what goes by

---

[9] This assumes that the applications or services they are utilizing actually attempt to use all capacity available on their connection. Video on the web today generally does not transfer at the highest speeds possible. (This includes video transmitted over TCP.)

[10] For a more complete discussion of Internet congestion and why the observance of congestion, in itself, is not an indication that a network is configured or managed inappropriately, see Bauer, Steve, David Clark, and William Lehr (2009), "The Evolution of Internet Congestion," paper prepared for 37th Research Conference on Communication, Information and Internet Policy (www.tprcweb.com), Arlington, VA, September 2009 (available at: http://people.csail.mit.edu/wlehr/Lehr-Papers_files/Bauer_Clark_Lehr_2009.pdf).

[11] The exception would be services like 100mbps broadband that is available in some countries where the service is actually defined by the line speed itself.

[12] DOCSIS is the Data Over Cable Service suite of international telecommunication standards that define interface requirements for offering broadband services over cable modems. The DOCSIS standards are developed by CableLabs, an industry research consortium (http://www.cablelabs.com/cablemodem/). Most of the cable systems currently in operation have implemented DOCSIS version 2.0, although a number of operators are presently rolling out DOCSIS 3.0 systems that support higher data rates and additional functionality.

[13] Lakshminarayanan, K., Padmanabhan, V. N., and Padhye, J. 2004. Bandwidth estimation in broadband access networks. In *Proceedings of the 4th ACM SIGCOMM Conference on internet Measurement*(Taormina, Sicily, Italy, October 25 - 27, 2004). IMC '04. ACM, New York, NY, 314-321.

the trade name of "Powerboost."[14] While Powerboost is advertised as a boost in the configured sending rate to a certain configured higher level, the Powerboost speeds often exceed even that advertised limit by non-trivial amounts.[15] This is important to recognize because some studies have attempted to infer the subscribed speed tier by assuming the maximum measured speed is less than or equal to 110% of the Powerboost level.[16] Tests we have conducted on our personal broadband connections have demonstrated this assumption does not hold.

So, if the configured peak or "up to" speed is not adequate for characterizing the speed of the network, how should speed be measured? Academics have long been interested in how to understand what a network is capable of supporting. The research literature focuses on a variety of different measurements one might want to make from the edge of a network. (We emphasize edge-based measurements because the research literature emphasizes how to infer or verify network characteristics with probing or measurements taken at the edges. Some of what one might want to know is known by the providers but may not be directly observable by third-parties.)

A distinction is generally made between the following properties of a network that one might want to measure.[17]

1. **Capacity** is a measure of the total traffic carrying capability of a link or path in a network. The end-to-end capacity is the minimum link capacity along a path. See Figure 2 (A) below. Link 3 in that figure has the most capacity but the end-to-end capacity would be determined by the capacity of link 2. The capacity is expressed as the amount of traffic the link can carry over a particular time interval (e.g., megabits per second).

2. **Available bandwidth**[18] is a measure of how much capacity is unused in a link or along a path. The available bandwidth along a path is the minimum available bandwidth of the set of links along a path. See Figure 2 (B) below. Link 1 has the most available bandwidth but the end-to-end available bandwidth is determined by link 2. One can look at a link utilization graph and observe that the total capacity was 100 mbps but the peak usage was 45 mbps and conclude that the available bandwidth, or spare capacity, was 55 mbps.

---

[14] See http://www.dslreports.com/faq/14520 (accessed April 22, 2010).

[15] For example one of the author's home cable network service includes a 12 mbps service plan with Powerboost up to 15 mbps. Measuring with tools like iperf and speedtest.net however we can regularly measure speeds in excess of 20 mbps for data transfers of 10 – 20 MB.

[16] This was part of the methodology employed by Comscore and relied upon by the FCC. http://www.netforecast.com/Reports/NFR5103_comScore_ISP_Speed_Test_Accuracy.pdf

[17] See for instance R.S.Prasad, M.Murray, C.Dovrolis, and K.C.Claffy. Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. IEEE Network, 2003.

[18] Network operators may be more likely to refer to "available bandwidth" as the "spare capacity" on a link or along a path. The "bulk transfer capacity" might be more likely to be discussed in terms of "throughput".

3. **Bulk transfer capacity** is a measure of the amount of data that can be transferred along a network path with a congestion aware transport protocol like TCP. See Figure 2 (C) below. The bottleneck capacity in that figure is link 2 but the bulk transfer capacity of a single TCP will be determined by the number of other competing TCP flows and the dynamics and settings of the TCP stacks on each end, the properties of the end systems, etc. A critical distinction is that the bulk transfer capacity is a measure of the data or "payload" throughput. Packet headers and data packets that are retransmitted are not counted toward the byte counts (and thus, such overhead may lower the observed speed).

   Also most of the academic literature and IETF RFCs define the bulk transfer capacity of a network as the achievable throughput of a *single* transport connection.[19] The assumption underlying this is that the transport protocol is well tuned and capable of fully utilizing network capacity if it is available. This is an assumption that does not hold for the default settings of many end user systems on the Internet today as we discuss further below.

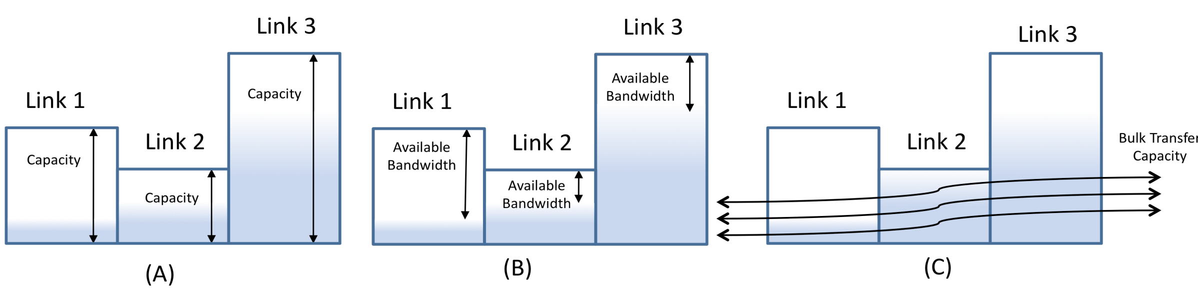

Figure 2: Network properties measurable with edge based measurement techniques. The research literature has focused in particular on measuring capacity and available bandwidth. Accurate measurement of the bulk transfer capacity along a path is a difficult task.

The "speed" that is most often reported and the focus of the tests we discuss in this paper most closely resembles this bulk transfer capacity. However, measures of capacity and available bandwidth remain vitally important for network operations, provisioning, and investment planning. For example, the available bandwidth, or spare capacity, in particular, is a main determinate of when capacity upgrades are initiated. While very important metrics, these are not as readily available and are not easily translated directly into an understanding of the end-user experience, and consequently, are of less interest to consumers.[20]

---

[19] RFC 3148 "A Framework for Defining Empirical Bulk Transfer Capacity Metrics," available at: http://tools.ietf.org/html/rfc3148).

[20] Industry analysts and large customers may consider issues like the availability of spare capacity to anticipate market supply and demand dynamics. Such information also has strategic value and so is more likely to be regarded as confidential by network operators.

For the purposes of this paper and our further discussion of the various speed metrics, we will focus on "speed" defined as the *bulk transfer capacity of a network measured with one or more congestion aware transport connections to some destination server*. Measuring even this seemingly narrow and precise notion of "speed" remains a significant challenge, as is often noted in the research literature.[21] For example, test measurements may differ because of differences in the number of congestion aware transport connections employed in the test, the location of the destination servers, and other factors that we address further below.

3.2. Average speed definitions

Producing summary statistics by averaging speed measurements over different populations (multiple users, long time periods, or across geographic regions) is a non-trivial task. It is common for discussion in the trade press and elsewhere to leave out the explanation for how summary speed statistics were calculated, instead simply stating the reported measurements represent the "average speed" for some population. It is as if one were to quote the national debt and fail to note that the quote was in dollars or yen. In the following sub-sections we highlight several notions of average speed computations that may be of interest and discuss some of the measurement issues that may arise.

3.2.1. Average access link speed

Assuming for the moment that there was a speed test that actually isolated and measured the speed of the broadband access network itself, free from factors outside a broadband providers' control,[22] then it is possible to calculate the "average" of such measurements in a variety of different ways, each with different potential interpretations.

For example, one could average the result of one user running a speed test at different times. Calculating this is important, but it is not clear that averaging these measurements gives the best indication of how congested the access network is. A different way of summarizing the same numbers would be the percentage of the day that the access speed was less than some fraction of the advertised speed. Such a measure would give some indication of how much congestion that user encounters.

Alternatively, one might be interested in computing an average that summarized the experience across users. For example, one could average the result of a number of users taking identical

---

[21] See for instance, R.S.Prasad, M.Murray, C.Dovrolis, and K.C.Claffy. Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. IEEE Network, 2003 which notes that "Several factors may influence TCP throughput, including transfer size, type of cross traffic (UDP or TCP), number of competing TCP connections, TCP socket buffer sizes at both sender and receiver sides, congestion along the reverse path, as well as the size of router buffers and capacity and load of each link in the network path. Variations in the specification and implementation of TCP, such as NewReno, Reno, or Tahoe, use of selective ACKs (SACKs) vs. cumulative ACKs, selection of the initial window size, and several other parameters also affect TCP throughput. [Citations removed]"

[22] Arguably this is strictly what broadband providers are advertising with their different speed tiers.

service offerings running a speed test. If this test were run at times when congestion was unlikely, then it would give some insight into the range of “baseline” performance different users see. This sort of average might be especially interesting for access based on DSL technology, since DSL speeds for different users will differ depending on circuit length and quality (see Figure 3). For other sorts of broadband access technologies (e.g., cable modem broadband), one would not expect to see a high degree of variation in “congestion-free” speed tests across users purchasing the same service tier.

If one averaged over congestion-free and congested times, then the results would be harder to interpret since it would conflate multiple factors (e.g., mixed line lengths that impact the achievable congestion-free speed for DSL and the impact of aggregate user traffic on shared facilities).

One could average the result of users with different providers or service offerings. Such a measurement might be useful for comparing broadband services across markets, but any such comparisons would need to be approached with caution because of the many factors that contribute to the differences (e.g., is it because of differences in the mix of technology platforms, consumer speed tier preferences, or traffic?).

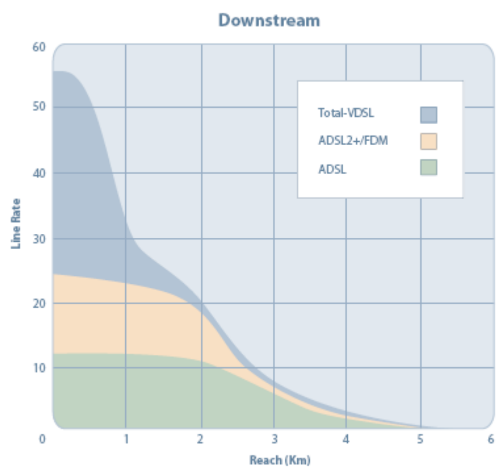

Figure 3: Line rate obtainable (Mbit/s) against corresponding line length (km) for ADSL, ADSL2+ and VDSL.[23]

3.2.2. Average end-to-end speed

The appropriate choice for summarizing the speed of end-to-end measurements is problematic. This is because end-to-end measurements are impacted by the end-user's configuration (e.g., is the test being run from a user machine on a home WiFi network that is connected to the ISPs

[23] Figure source: http://en.wikipedia.org/wiki/ITU_G.992.1. Also see the same page for a corresponding graph that displays the line rate obtainable (Mbit/s) against corresponding line attenuation (dB).

broadband service), server performance, TCP tuning, and path performance (e.g., how is traffic routed beyond the access ISP to the server?).

One could record and average all the speed tests at a given test server. If this data were gathered at the server, an appropriately instrumented server could distinguish those samples that were limited by server performance and those that were not (thereby isolating at least one important source of variability).[24] If the average included only those samples where the server was not the bottleneck, then the resulting average might give some hint as to the range of throughputs different users actually experience when using the Internet, taking all the host and network effects into account. If the average were computed for a single user access connection, the measurement might provide a hint as to the level of overall user satisfaction, but even in this case, it would not allow one to determine whether any service problems were due to the access ISP's network or due to bottlenecks elsewhere along the path to the server.

Alternatively, one could average all the speed tests made by one client to different servers. These tests would hold the behavior of the client (e.g. the TCP tuning parameters) and the access link constant, and see what variation arises from all the other factors, most obviously the round trip delay and bandwidth of the paths to the different test servers, as well as the servers themselves. Such an average would give that user some sense of how well his ISP is connected to the rest of the Internet (e.g. are there bottlenecks to certain other regions). However, this average would be difficult to interpret since many implementations of TCP show substantial performance variation based solely on round trip delay. This average might fail to tell the user anything about path bottlenecks.

As the preceding discussion makes clear, the choice of population over which an average speed is computed has a huge effect on how the measurement may be interpreted. Since summary statistics are both necessary and will be used, we caution readers to try and be clear regarding how average estimates are labeled (and the methods for their calculation explained) so that they may be interpreted appropriately.

4. Speed measurements

In this section we turn to a discussion of some of the popular speed testing services that are in current use. While we recognized that the choice of what is measured makes a big difference, we were surprised by the significant variability across what originally seemed to be attempts to measure the same thing. In the following sub-sections, we first present some of the data that stimulated our interest and then the results of our more in-depth analysis of what underlie the measurement discrepancies for several of the more common testing sites. In light of the attention

[24] The NDT test (one of the speed tests on broadband.gov) collects information from the TCP stack adequate for this process. However, we don’t believe they are currently taking this or other non-network bottlenecks into consideration when looking at the data they have collected.

that these data have had in the press and broadband policy debates, a closer understanding of these data is important for their appropriate interpretation.

4.1. Puzzle of broadband speed measurement variability

In Table 1 below, the average US download speeds are reported from Speedtest/Ookla, Comscore, Akamai, and Youtube. The variation in measured speeds at smaller geographic regions is even greater. At first glance, the Speedtest/Ookla measurements in this table (and generally) look like they might be outliers as they consistently report far higher speeds than the others. Results like these induced us to ask questions like: Did the tighter range of numbers from other organizations represent confirming evidence that they were actually more representative of the speed of the broadband access networks? And critically, did each organization have a sufficiently similar notion of speed to make these results actually comparable?

Reporting site	Reported US download speeds	Reported MA download speeds
Speedtest/Oooka	7.93 mbps[25]	11.54 mbps[26]
Comscore	3 – 4 mbps[27]	(none reported)
Akamai	3.9 mbps[28]	4.9 mbps[29]
Youtube	4.21 mbps[30]	7.39 mbps[31]

Table 1: Download speeds reported by various organizations

We were also intrigued when we ran the individual speed tests that were hosted on the FCC's Broadband.gov website.[32] Sequential tests alternated between employing the Speedtest/Ookla test engine and the NDT test hosted on the Measurement-labs infrastructure.[33] Figure 4 below presents screen captures of eight sequential tests run from a cable modem in the Boston area with a 12 mbps (Powerboost to 15 mbps) service tier. The upload speeds are fairly consistent across all eight tests. The download speeds for the Speedtest/Ookla are also fairly consistent at 16 mbps. The download speeds however for the NDT tests vary between 2.6 mbps and 14.3 mbps.

[25] http://speedtest.net/global.php (accessed 5/1/2010)

[26] http://speedtest.net/global.php#0,1,1 (accessed 5/1/2010)

[27] See U.S. National Broadband Plan http://download.broadband.gov/plan/national-broadband-plan.pdf

[28] Q-3 2009. http://www.akamai.com/stateoftheinternet/

[29] Akamai Comments on NBP PN24 12-14-2009.pdf

[30] http://youtube.com/my_speed (accessed 5/1/2010)

[31] http://youtube.com/my_speed (accessed 5/1/2010)

[32] http://www.broadband.gov/qualitytest/about/ (accessed 5/7/2010).

[33] http://www.measurementlab.net/ (accessed 5/1/2010)

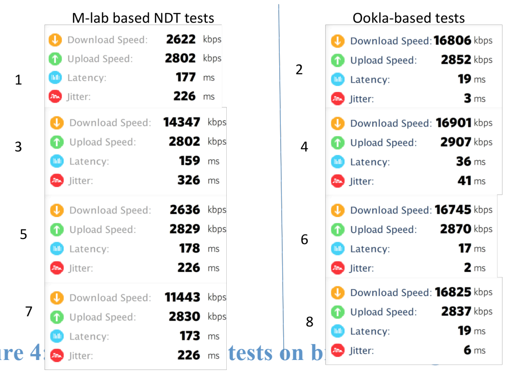

Figure 4: tests on b

Did we just get unlucky and experience access network congestion on the low NDT tests or does something else account for the variation? The first clue to an explanation comes from noticing the significantly different latencies measured across the different tests. The Speedtest/Ookla measurements were all done to a nearby server less than 20 ms away, while the NDT tests all have latencies greater than 150 ms. After looking at the packet traces for these tests, it became clear that the NDT tests all happened to connect to one of the measurement lab servers in Amsterdam.[34]

The Ookla/Speedtest measurements were much more inline with what we have previously measured for that broadband connection using testing tools like iperf[35] and applications like wget[36] for transfers and tests between MIT (which we know was not a bottleneck because we have detailed information about how it is configured) and this particular home network. This suggested more generally that the Speedtest/Ookla numbers might not be an outlier in Table 1 above. Data such as these induced us to engage in in-depth detective work to uncover some of the core differences across the testing sites.

4.2. Site-specific measurement differences

In the following subsections, we summarize how a number of the best-known speed testing tools/sites measure speed and highlight some of the issues that arise and may lead one to prefer one approach over another. Some of the key differentiators include:

- How close is the test server likely to be to the user (i.e. latency)?
- How many TCP connections are used to run the test?

---

[34] There are m-lab servers in the United States, in particular, one close to Boston is in NYC. The remote server selected is based upon both test server load and proximity to the client. The Amsterdam server just happens to service some of the system load from the Boston area.

[35] See http://iperf.sourceforge.net/

[36] See http://www.gnu.org/software/wget/

- How are sample measurements filtered?

These differences can have a significant impact on the resulting speed that is measured. For instance a test run between a client and server with an average round trip time of 30 ms versus the identical participants running the same test with an average round trip time of 300 ms can result in measured speeds that differ by a factor of ten. This is not a paper about potentially inconsequential measurement differences of 5%.

We begin by discussing ComScore, which is important because it has been relied upon by the FCC, but is different from the other speed data sources we consider because our analysis is hampered by a lack of primary data.

In the case of the other sources for speed measurement data, the basis for our work was the public documentation of methodologies, helpful private conversations with the testing organizations, and manual inspection of packet traces taken when speeds were being measured. While this was enjoyable detective work, it is clear that any organization that makes speed testing tools or results publically available needs to better document their testing methodology.

Beyond better documentation, we would like to see speed testing organizations post a packet trace of a test and the corresponding measurements they would produce for that trace. (Ideally this would be a set of traces and corresponding measurements.) Such documentation with trace-based samples would have greatly assisted us in better understanding and validating the measurement methodologies.

4.2.1. ComScore

ComScore is a marketing research company[37] that among its services provides data on broadband speeds based on its own test methodology and measurement infrastructure. Comscore is of special interest because their data has been used by the FCC and is cited both in reports they issued in September 2009 documenting the current state of broadband in the United States[38] and in the National Broadband Plan.[39]

Among the testing data sources we consider here, it is the only one for which we were not able to analyze primary source data. We had no access to documentation, no ability to manually test their software, and no chance to examine packet traces of their individual tests. However, a report by Peter Sevcik (NetForecast, March 2010) on the ComScore data that was submitted to the FCC identified a number of methodological and other issues that raise concerns that we share.[40] The methodological description he had to work from was evidently somewhat imprecise

---

[37] See http://www.comscore.com/.

[38] See http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-293742A1.pdf (accessed 5/7/2010).

[39] See http://www.broadband.gov/download-plan (accessed 5/7/2010)

[40] ComScore ISP Speed Test Accuracy http://fjallfoss.fcc.gov/ecfs/document/view?id=7020397512.

or inaccurate.[41] However, a number of the concerns he raises appear to be valid and potentially significant sources of measurement error. In particular and as we explain further below, because ComScore bases its measurements on a single TCP connection, it will tend to significantly under-state the achievable speed of the broadband network service. The receiver often limits the speed by advertising to the sender that they do not have adequate buffer space to handle data transferred more quickly (or more technically the sender is limited by the receiver window instead of the congestion window.)

Another significant potential source of error identified in Mr. Sevcik's report is the improper inferences about users' speed tiers -- particularly on cable networks which employ Powerboost. ComScore assumes that the maximum speed they measure will be at most 10% above the advertised Powerboost level. We have packet traces available at our wiki that demonstrate that this is not a sound assumption.[42]

4.2.2. Speedtest/Ookla

The Speedtest/Ookla data is based on the test software and measurement infrastructure developed by Ookla Net Metrics, a developer and vendor of networking testing applications.[43] In addition to their test engine, Ookla also maintains a free testing website, speedtest.net. The speedtest.net site enables visitors to measure their performance to any public Ookla test server (hosted by partner organizations) around the globe -- hundreds of sites around the world currently run it. Over 1.4 billion tests have been run using the Speedtest/Ookla engine since it began operating in 2006.[44]

The Speedtest/Ookla measurement methodology and summary statistics consistently report the highest speeds for both individual results and aggregates at the provider, state, and country levels. Each test consists of measuring the upload and download "speeds" and latency between the client and server.

4.2.2.1. Methodology Description

The Speedtest/Ookla testing engine is the measurement methodology that was the easiest to investigate for a number of reasons. The Ookla company had posted a public description of their

[41] A response from comScore was published as a comment here http://www.networkworld.com/community/node/59354#comment-240274. In this comment they provide a description of exactly how they calculate the speeds addressing a surprisingly common calculation error that arises because mega has slightly different meanings for file sizes than it does for network bit rates. A 1MB file is actually 1048576 bytes not 1,000,000 bytes.

[42] http://mitas.csail.mit.edu/wiki

[43] http://www.ookla.com/

[44] http://speedtest.net

test engine's general methodology which we reproduce in the table below.[45] They also answered our email and phone queries and updated their test description to clarify a number of steps that we found unclear. Perhaps most important to furthering our understanding though was our ability to actually run the test while capturing packet traces. This enabled us to investigate the packet traces to validate (according to their stated methodology) that the reported speeds were consistent with the traces. Ookla also offers a free trial server that we installed and ran on MIT's campus.[46] This enabled us to experiment with various configuration options and capture packets on both the server and client side.

Download test	Upload test
1. Small binary files are downloaded from the web server to the client to estimate the connection speed	1. A small amount of random data is generated in the client and sent to the web server to estimate the connection speed
2. Based on this result, one of several file sizes is selected to use for the real download test	2. Based on this result, an appropriately sized set of randomly generated data is selected for upload
3. The test is performed with cache prevention via random strings appended to each download	3. The upload test is then performed in configurable chunk sizes (pushed to a server-side script via a POST)
4. Up to 8 parallel HTTP threads can be used for the test	**4. The test can be done using up to 8 parallel HTTP threads (configurable)**
5. Throughput samples are received at up to 30 times per second	**5. Chunks are sorted by speed, and the fastest half is averaged to eliminate anomalies and determine the result**
6. These samples are then aggregated into 20 slices (each being 5% of the samples)	
7. The fastest 10% and slowest 30% of the slices are then discarded	
8. The remaining slices are averaged together to determine the final result.	

We have highlighted in bold the steps that have the most influence on the speeds that are reported by this methodology. In particular all Speedtest/Ookla sites that we tested employed at least two (four also being common) TCP connections for tests in each direction. In addition to being far more likely to run the test to a nearby server[47], the number of TCP connections is the biggest differentiator of this test and the other methodologies we consider. The combined throughput of all TCP connections is the input into the next critical step in their methodology where they filter the throughput samples they take throughout the test period.

[45] http://wiki.ookla.com/test_flow

[46] http://www.ookla.com/trial.php

[47] The default server is the one geographically closest to the client.

Before examining the filtering step, we address why Ookla employs multiple TCP connections and whether that is representative of usage patterns on the Internet today. The answer to this is quite simple: a single TCP connection is unable to saturate an uncongested link with many of the operating systems still very common on the Internet (the data we examine in the NDT section below conclusively demonstrates this). The bottleneck becomes the client system instead of the broadband access network. By employing multiple TCP connections they don't harm the performance of tests running from machines that would be capable of saturating a single TCP connection, and they gain the benefit of moving the likely performance bottleneck back into the network (assuming that is what the measurements are trying to understand).

Multiple simultaneous TCP connections are representative of many typical applications on the Internet today. Web browsers like Firefox, Microsoft's Internet Explorer 8, and Google's Chrome open four to six simultaneous connections per domain. Coupled with the common practice of websites using two or more sub-domains, the number of simultaneous TCP connections can easily be four to eight.[48] Other applications like peer-to-peer traffic similarly open multiple simultaneous TCP connections. (There are exceptions to this -- web-based video traffic tends to be carried on a single TCP connection and large file transfers also employ a single TCP connection.) So employing multiple TCP connections both increases the odds that the test will measure a bottleneck in the network and is a representative test for many common usage patterns.

The next important differentiator for the Speedtest/Ookla methodology relates to how they filter the test results. Ookla's results are based on disregarding the fastest 10% and slowest 30% of the measurements collected while the TCP connections were in progress. This is important because they deliberately do not take the total data transferred divided by the total time – a measure employed by the ComScore and NDT methodologies (but not the Akamai or Youtube methodologies). Filtering the sub-samples to retain and average the 60% has been referred to as a measure of "peak" speed.[49] We disagree with this characterization. Peak can be defined in a number of different ways in networking (i.e. the highest sampled interval, the 95th percentile sample, etc.), but the average of the 30th percentile though the 90th percentile is not a typical measure of peak.

The average of the 30th percentile though the 90th of the samples is most likely higher than the overall average. So what is the possible justification for excluding time intervals with lower speed measurements? The answer to this lies in understanding the complicated dynamics of the TCP protocol. TCP is continually probing the network to find the appropriate sending rate based upon the capacity and current congestion level in the network. It has complex algorithms that

---

[48] Google makes exactly this observation in arguing for an increase in TCP's initial congestion window as well. They argue that the *effective* initial congestion window for browsers today is 80-120 segments for applications like Google Maps (instead of 3 segments for the canonical single TCP connection). http://code.google.com/speed/articles/tcp_initcwnd_paper.pdf

[49] "Consumer Broadband Test Update", blog posting on http://blog.broadband.gov/?entryId=292153

have continued to evolve for recovering from lost packets and ramping back up once the loss epoch has passed.[50]

Figure 5 presents an annotated Time/Sequence graph produced by the packet capture program Wireshark that graphically shows the number of packets sent at different time increments and provides an illustration of some of this complexity. The data is based on capturing one of the NDT tests mentioned above from the Boston-based client to the server in Amsterdam. The speed can be inferred by the rise over the run of the dark line at different points. (The lighter line above it represents the receiver buffer capacity, or receiver window. It is not a bottleneck during this particular transfer but often is in others.) The question of interest is which slopes to measure in reporting the “speed” of this transfer? The first packet wasn’t dropped by the network until about 3.5 seconds into the connection. TCP was able to recover most of the lost packets through about 5 seconds but then took a considerable amount of time slowly recovering through to about 10 seconds where it finally starts ramping up its sending rate again. Even then it very conservatively estimates the network capacity and grows its sending rate more slowly than it did during the initial 5 seconds of the transfer.

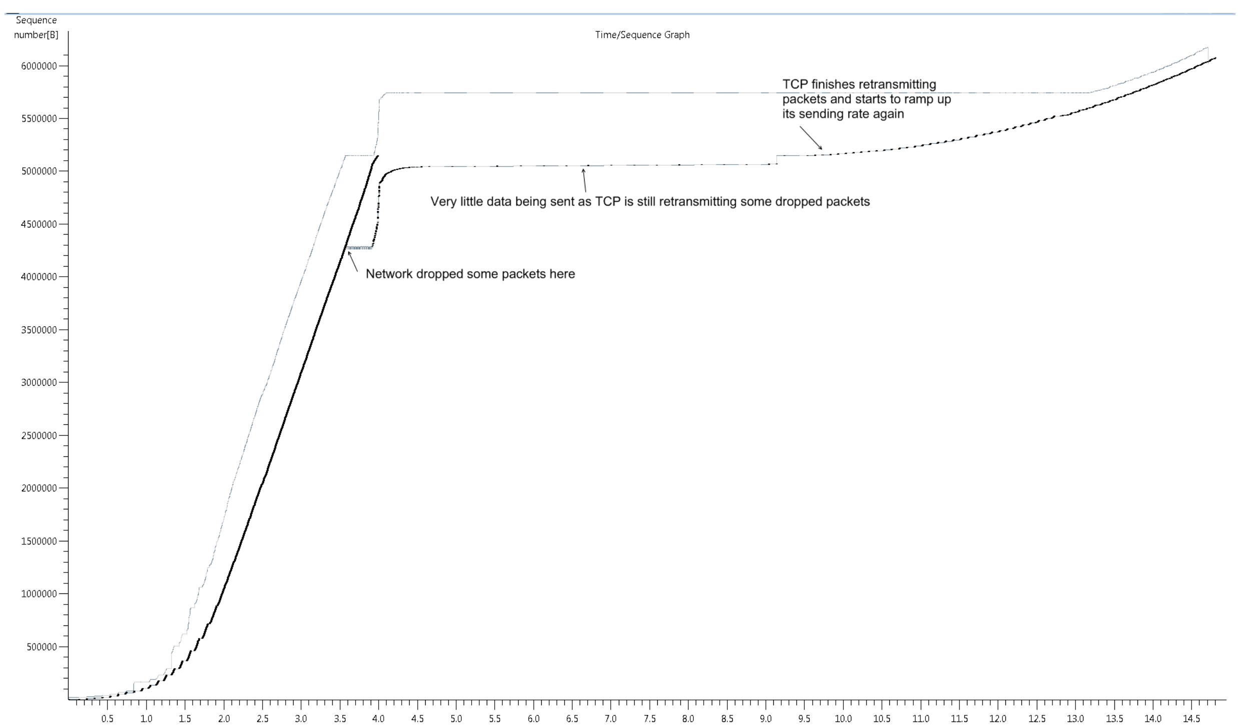

Figure 5: Time/Sequence number plot of a TCP connection that takes a very long to recover from a lost window of traffic.

One cannot definitively infer from this plot when the access network was actually a bottleneck (if indeed it was at all -- the more likely bottleneck in this particular case was the transatlantic

[50] See Bauer, Clark and Lehr (2009) note 10 *supra*.

link.) But if we assumed for the moment that the access network was responsible for the lost packets starting at 5 seconds, is it accurate to take the amount of traffic TCP transferred during its recovery periods as representative of what the connection was capable of carrying? It is indeed possible the access network was what constrained the data rate but the fact that no other packets were lost during this recovery period strongly suggests that more carrying capacity was available. TCP is deliberatively conservative in recovering from packet losses.

But why does Ookla adopt 60% as opposed to some other fraction of the samples they measure? If one accepts the argument that not all samples of throughput during a TCP connection are representative of the actual carrying capacity of the network, which samples are representative? This is impossible to determine definitively from edge-based measurements. From talking with Ookla, the decision to retain 60% was an engineering decision based upon a non-trivial but not necessarily systematic testing. (It would be interesting if this were a configurable parameter of their test engine then one could easily run a systematic study where all elements of the network path were carefully controlled. Based upon the testing and packet traces we did conduct, retaining 60% does not appear to be an unreasonable measure.)

One of the configurable parameters of the Ookla engine is the test length – the default for both the upstream and downstream direction is ten seconds. This is a variable that does have an effect on the speeds that are measured particularly on network connections that employ a Powerboost like technology where the beginning of long transfers will be faster than the long run transfer speeds. (See Figure 6 where there is a definite knee in the plot bending to a lower speed after the Powerboost period expires.)

This raises the question, how long should one measure the throughput to determine the speed of the network? For web content, Google posts the statistical information derived from their web crawling that suggests that at least 90 percent of pages will easily load with such a Powerboost window given the total amount of KB per page.[51] So primarily measuring the Powerboost period seems acceptable for web-like usage patterns today. Similarly measuring the Powerboost period is likely acceptable as representative of the initial buffering period employed by video sites like Youtube that download an initial buffer quickly and then download the rest of the file more slowly as the user actively watches it. The iperf test software also adopts the default time of 10 seconds for its tests. This is, of course, not to say that 10 seconds is definitively the right time interval to measure. This has to be selected based upon what constitutes the workload one is attempting to approximate with the speed measurement. This is a bit like asking how fast a car is and wanting different sorts of answers if one is thinking of drag racing over a 100 yards or a cross-country rally.

---

[51] http://code.google.com/speed/articles/web-metrics.html

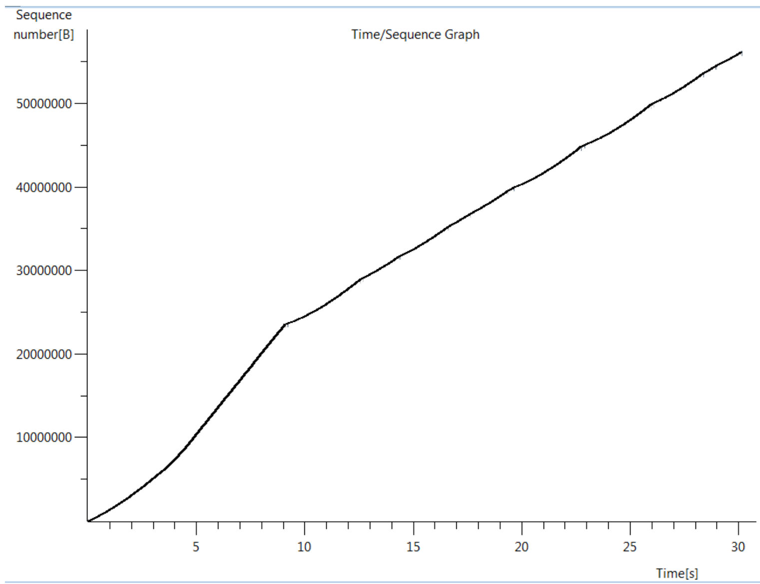

Figure 6: Time/Sequence plot of a single TCP transfer showing a definite knee in the throughput (around ~20MB, 9 seconds) after the Powerboost period.

While performance tests can be run to any Speedtest/Ookla server around the globe, the Speedtest.net site identifies the geographically closest server to the client. Most users likely accept this default test server. This is important because latency to the test server can have a tremendous impact on these (and all other methodologies') results. Geographically further sites are usually correlated with lower speeds.[52]

4.2.2.2. Implementation

The test is run between a Flash based applet embedded in a webpage and server side hosted on a web server. The default server to test is the one in closest geographic proximity to the client. All the calculation of download and upload speeds and latency is performed on the client side. Each of the simultaneous TCP transfers are HTTP requests to a web server. Testing servers are located throughout the world run by numerous organizations. Test results are reported back to an Ookla server that aggregates results from clients.

Since the client is software based, the operating system and computer hardware of the user are out of the control of the test. The local area network is also out of the control of the test procedure -- other traffic could be on the local home network from other computers (e.g., someone is watching a movie while someone else is executing the test, resulting in a lower speed test result) or the computer running the test itself (e.g., the computer is running an on-line backup

[52] This is not always the case. For a cable modem we test frequently in Boston, the speed to NYC is generally higher as the route to the default test server in Boston happens to travel down to NYC first then back up to the other provider in Boston.

in the background). Tests run from clients connected via home WiFi (802.11) networks may be experiencing WiFi bottlenecks that would result in slower speed results.

4.2.2.3. Speedtest/Ookla public data

Summary statistics from the over 1.4 billion Speedtest/Ookla tests have been publically available for a long time on the Speedtest.net website.[53] Speeds are reported for both providers and different geographic areas like city, state, and country. The method for calculating these aggregates is to calculate the 95th percentile speed in each direction for every unique IP address that has been tested at Speedtest.net. If fewer than twenty tests have been taken the highest speed measured by an ISP is taken. These numbers are then averaged together for each geographic level (both overall and per ISP). They currently include all tests ever done using their tool in this calculation. Data is from August 2006. Their methodology has changed somewhat over time as they didn't start using multiple threads until August 2007.

A new analysis of the Ookla/Speedtest data was made available in 2010 on a NetIndex.com website.[54] The key difference in this dataset compared to the one above is that the reported speed is a rolling average of the throughput measured over 30 days where the mean geographic distance between the client and the server is less than 300 miles. This is a more timely view into the data and a view that is more likely to be representative of the broadband access link speeds. Accordingly these speeds will likely be slightly higher than the other summary views.

Ookla/Speedtest source data is available for research purposes. Source data that provides completely anonymous daily index values going back to January of 2008 for the geographic locations in the NetIndex are available publically. The raw data is available to academic researchers. At this point in time, both of these versions of the data are available for free. Making these data sets publically available greatly assists academic research and policy analysis. The publically available M-labs NDT test data is a similarly rich set of raw data. We wish other organizations publicly reporting speed measurements would likewise make available the underlying raw data.

4.2.3. Akamai Speed Reports

Akamai is a company that provides content and application delivery services over the Internet. They operate a content delivery network (CDN) that consists of tens of thousands of servers in many of the largest 1000 networks on the Internet. Their delivery network serves content for a wide selection of web sites and thus observe web traffic from a wide selection of end users.[55] For both billing and operational reasons they keep fairly extensive logs of clients that connect to their servers to download content. They have analyzed these logs and produced a "State of the

---

[53] See http://www.speedtest.net/global.php

[54] See http://www.netindex.com/

[55] Selected customer list is available at: http://www.akamai.com/html/customers/customer_list.html

Internet" report quarterly since the first quarter of 2008. In addition to reporting on connection speeds, they also present data on attack traffic, Internet penetration and broadband adoption, and general trends seen in the data. These reports, and similar FCC filings, present the "average connection speeds" for different geographic areas. This measures the download speeds only; upload speeds and latencies are not reported.[56]

While end users' normal web activity naturally contributes to the Akamai data set as users visit sites that serve content through Akamai, users cannot manually test their individual speeds. So we were not able to gather packet traces and compare them to the speeds that were measured. Akamai, however, was very responsive to our inquiries for more information.

Akamai states that the connection speeds they report are measurements of throughput between clients and their Akamai network of servers. They are not intended as a measure of the access link speed. In particular, when asked about this we were pointed to one of their FCC filings which notes that "Akamai's measurements do not conform to the reporting categories of the FCC's Form 477 (e.g., speed tiers) but rather are based on measurements that Akamai's customers have found useful."[57] What they log and analyze is driven by the needs of their paying customers. We are pleased they choose to analyze and expose their data to contribute to a better public understanding of the Internet.

4.2.3.1. Methodology Description

While the methodology for producing the raw individual speed numbers is not publically documented, Akamai informed us that each HTTP GET request to their infrastructure produces a log entry that includes, among other information, the requesting client IP address, the bytes transferred and the time duration. From this one can calculate a speed per transfer for a client. Multiple HTTP requests can be served sequentially over the same TCP connection. So each individual speed measurement represents some part of the total TCP duration. A benefit of this methodology is that each request does not need to go through the TCP startup phase – subsequent requests benefit from congestion state learned by TCP from the preceding requests.

Multiple TCP connections from a single client can be opened simultaneously to the Akamai servers. Each HTTP request over each connection produces its own speed measurement. Throughput rates from simultaneous connections per client are not added like in the Ookla methodology. Because connections to Akamai occur by definition when the user or their applications are active, simultaneous connections to other non-Akamai servers are also possible. If the access network were the bottleneck, all of these effects would tend to make the resulting measurements systematically lower than the access link speed even if there was no congestion on the access link.

---

[56] We would not expect upload speeds as the Akamai infrastructure is setup to serve content to customers.

[57] See http://fjallfoss.fcc.gov/ecfs/document/view?id=7020352915

Akamai maps clients to servers based upon both network proximity and load on the Akamai servers. In general this will result in clients connecting to nearby servers with low latency (i.e. the best conditions for measuring TCP throughput.) However, if part of the Akamai infrastructure does happen to be loaded, clients will be directed to other nearby servers (still likely close, but not as ideal for measuring throughput.)

4.2.3.2. Implementation

The exact details of how Akamai calculates the speed for an individual HTTP transfer is proprietary. But in general it is obviously a sender side implementation. For performance, deployment, and maintenance reasons, the estimate of the duration per transfer is almost assuredly an application level estimate. There are a variety of ways this could be done with a high degree of accuracy. However, it is non-trivial because the server application writes data to a socket and does not know exactly when the receiving side acknowledges that data.[58]

The file sizes that speed estimates are produced for in the Akamai data will vary based upon the distribution of content Akamai happens to serve. While some other speed tests suffer from testing the duration of time to transfer a file that is inadequately small and doesn't allow TCP to grow its sending rate to fully utilize the available network capacity, this is less likely to be the case in the Akamai data since their speed estimates are calculated for each HTTP transfer that are part of a single TCP connection. The earlier transfers will have likely grown the TCP sending rate. However, if a small file is requested at the beginning of a TCP connection, or is the only request, the estimate for its transfer speed may have been limited by the rate at which TCP increases its sending rate, not any actual bottlenecks in the network. These effects would result in calculating lower connection speeds. Akamai most likely filters some of their log entries in producing their connection speed estimates to ameliorate some of these effects but we do not know the exact procedure.

Akamai is measuring regular network activity. So the application, operating system and computer hardware of the clients connecting to the Akamai server are out of the control of the measurement methodology. As is de facto the case in the Akamai test, the computer making the request is engaged in application and, often, other network activity. These effects strictly reduce the speeds that are measured. The question is how large are these effects? The answer to this is that it depends upon a great many factors that are not directly knowable. What Akamai measures is assuredly correlated to the access network speed but should not be taken as an absolute measure of the performance characteristics of broadband access networks.

[58] The Ookla and NDT tests face a similar challenge. Again in both cases we assume it is an accurate estimate.

4.2.3.3. Public Data

The aggregated data that Akamai publishes is calculated according to the following procedure taken from their FCC filing:

"[F]or each individual IP address that connects to an Akamai server, daily mean transmission speeds are collected and used to produce 7-day means. Four consecutive 7-day means are used to produce a 4-week mean. Then three consecutive 4-week 3 means are used to produce a quarterly mean speed. Then the quarterly mean speeds for all IP addresses are aggregated and averaged."

Akamai has very good geo-location techniques as this is a core part of how they map users to nearby content servers. So their ability to correctly identify connections as originating from a particular area is assuredly quite good. They aggregate data across all clients connecting to their servers, which include universities, businesses, and residential networks. Each of these types has fairly different connection profiles. It is unclear how the distribution of types of clients varies across the different regions that Akamai reports on or what fraction each client population contributes to the averages reported. As others have noted, when Akamai reports the speeds for a given city it seems likely that the average speed is affected by the presence of universities with high speed Internet connections.[59] The point is that the Akamai data characterizes web traffic as a whole, and is not limited to residential broadband access.

4.2.4. Youtube Speed numbers

Youtube is the video sharing website owned by Google. In early 2010 they began allowing users to see detailed estimates of the speed Google estimated for their individual connection, provider, and regions when viewing Youtube videos. Any user can see their data by visiting http://youtube.com/my_speed. A sample is shown in Figure 7 below for a cable modem user with a 12mbps connection that has Powerboost to 15mbps.

[59] See http://www.digitalsociety.org/2010/04/the-akamai-speed-report-is-skewed-too/

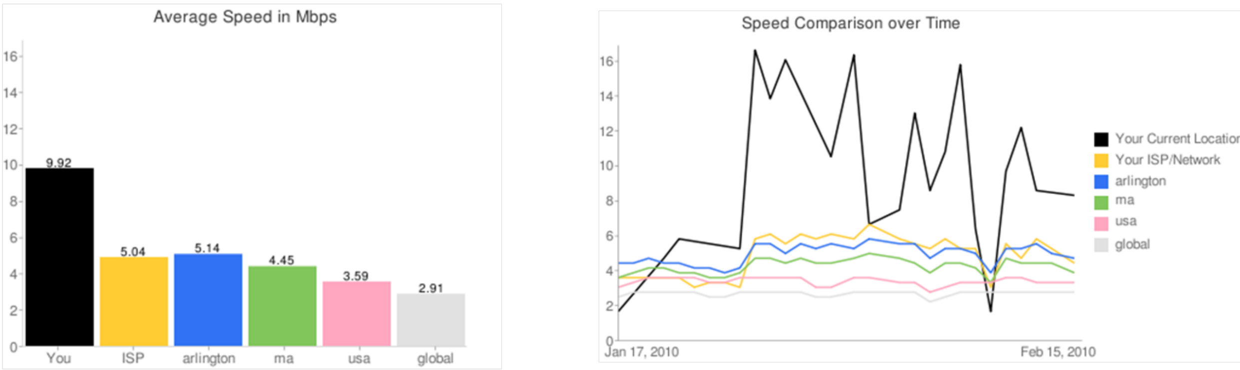

Figure 7: http://youtube.com/my_speed

This is a fascinating move by Youtube. Youtube clearly would benefit from higher speeds between their servers and their end users as they would be able to transport videos that are encoded at higher qualities. They are explicitly inviting users to compare their speeds to both competitors ISPs and to speeds in other geographic regions.

We wonder if other content providers (video and otherwise) might soon start exposing similar performance details for their user populations. Though we have concerns explained below with how some are likely to interpret the Youtube measurements, we overall think this is a very healthy and innovative sign for the Internet ecosystem. More systematic and public measurements about Internet performance will help drive investigations into performance bottlenecks and motivate individuals and companies to seek the highest performance level for a given budget.

4.2.4.1. Methodology

In the FAQ Google provides very general information about the methodology they employ for estimating connection speeds. They note that

> *We first compute the bandwidth of almost every YouTube video played by considering the amount of data sent to and acknowledged by the user's machine in a given time period. Small video responses are excluded as they can add noise to our bandwidth calculation. This bandwidth estimate of every video played is associated with a VISITOR_INFO1_LIVE cookie and with the IP address that requested the video, but NOT with a YouTube user name.*
>
> *For the daily speed number as shown in the YouTube speed page we average the bandwidth for all videos played from a given cookie and IP address across the course of the day. If a single laptop user uses multiple network connections and watches videos regularly from these, they could see different speed measurements and historical data based on which connection they are using.*
>
> *The measurements for geographical regions are computed by averaging the daily averages of all users (based on VISITOR_INFO1_LIVE cookie) who played videos*

in the same geographical region. The measurements for the Internet Service Provider (ISP) are got by averaging the daily averages of all users (again based on VISITOR_INFO1_LIVE cookie) who use the same ISP and are in the same geographic region.

We were not able to determine exactly what parts of the video stream the speed estimates were taken from. It cannot possibly be the total data transferred divided by the transfer time of the TCP connection. The initial buffering period loads the video as quickly as possible; additional parts of the video then are transferred at fixed intervals as the video is watched. So there are considerable gaps where data is deliberately not being tranferred. The speed estimates therefore must be taken during the initial buffering period or during these data bursts. (See the Time/Sequence plot of a packet capture in Figure 8 with an embedded window zoomed in on the peroidic data bursts.)

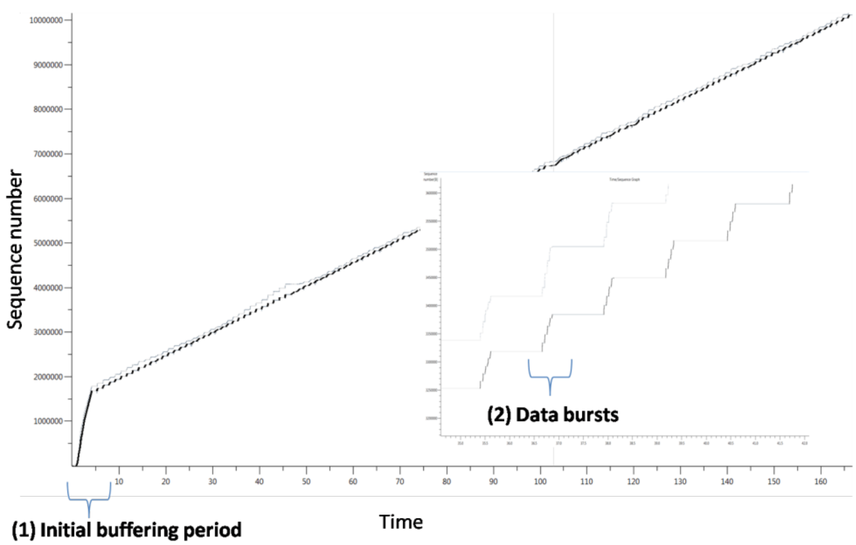

Figure 8: Time/Sequence plot of a packet capture for a single Youtube video. The initial buffering period loads video as quickly as possible. Additional parts of the video then are transferred at fixed intervals as the video is watched. The speed estimates must be taken during the initial buffering period or during the data bursts.

What puzzles us about the Youtube methodology is the results that we get from viewing the speed report from MIT's campus. MIT is generally considered to have very good local area networks, both wireless and wired (gigabit LANs in most places). Traceroutes from MIT to Youtube reveal that we peer with them at an exchange point in NYC. MIT network administrators have told us that our connection into that exchange is a gigabit Ethernet link that has an average load of 30mbps that tends to peak at 50mbps. In other words, our connection to the exchange doesn't appear to be the bottleneck.

However, the speed numbers we captured in March (see figure below) from two computers on the same floor of our lab at MIT were both remarkably different and much lower than we expected. (The misidentification of the 18/8 network as a non-Cambridge network have since been corrected.) Both 18/8 and 128.30/16 have a mixture of wired and wireless computers, both these computers happened to be on wired networks. MIT users certainly have capacity to transfer data to many different places at speeds well in excess of 10 mbps. But what surprises us is the amount of variation seen in the measurements across the month. MIT's capacity did not vary over this time period and the network load did not vary by an amount sufficient to cause these variations in speeds either.

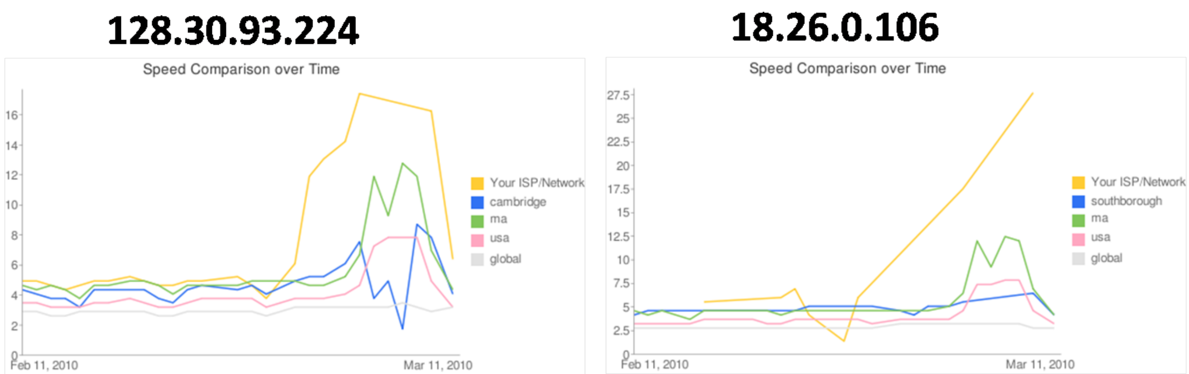

Figure 9: Speed numbers from http://youtube.com/my_speed

Google assuredly did not make a simple calculation error. So why did they measure the speeds they did? There are a number of possible explanations for this. They may have changed their measurement methodology over time as the line "*Small video responses are excluded as they can add noise to our bandwidth calculation"* was added to the FAQ after the initial release of their site. We have no indication of whether or not this occurred. They would likely need to change this lower size limit over time as connection speeds rose. This gives one possible clue to why MIT's numbers may vary so much. Measurements of MIT's speed may be very noisy because our connections are so fast. Since Youtube must be taking their measurements from parts of the TCP connection, potentially TCP doesn't have time to ramp up fast enough to fully exploit the available capacity or their ability to accurately time very fast connections may be limited.

Their methodology simply may be much more tuned for measuring the experience of typical residential broadband customers. The Speedtest/Ookla methodology for instance has various configuration parameters (like the file sizes and number of TCP connections) exactly for tuning the test methodology to better measure certain expected ranges of connection speeds. On the residential broadband networks we have tested, the speeds reported by Youtube are consistently lower than the speeds we are able to measure with either the Speedtest/Ookla site or the manual testing we do with tools such as iperf that we mentioned above.

This suggests to us that while the speed numbers reported by Youtube likely are correlated with the broadband access network performance they should not be interpreted as an absolute measure of performance. In particular the performance bottleneck may not be the broadband access provider itself. The Youtube speed data is presented in such a way that seems easy to misinterpret as directly attributing the measured speeds to the access ISP itself (i.e. labeling the speed results with "Your ISP/network" instead of the end-to-end path that includes the user, their ISP, the upstream ISPs, and finally Youtube itself). Any of these can be the performance bottleneck. The FAQ notes[60] that "[t]here are many factors that affect the video speed. Some of the factors that impact the video speed are the Internet Service Provider you are using, the distance from your computer to Google servers, the computer you are using, other devices in your network such as other computers and Internet connected appliances. etc." Most customers and policy makers though may not notice these caveats and incorrectly assume the numbers can be compared to the speed tier they selected.

4.2.5. M-Labs Network Diagnostic Test (NDT)

In addition to the Ookla/Speedtest, the Network Diagnostic Test (NDT) is the other major test currently running on the broadband.gov website. Before its use by the FCC, this test was being employed by the scientific community to understand performance on their high capacity research networks.[61] It then became one of the many network tests hosted on the Measurement Lab (M-lab) server infrastructure.[62] It has also recently been integrated into application software from Bittorrent.[63] In general it is a widely used test for measuring TCP throughput in both the upload and download direction.

Its importance as a testing tool lies in the extensive logs of TCP state variables it collects. It leverages an instrumented TCP stack[64] to track the evolution of the connection state over time. (It also collects a packet trace of every test -- one could not hope for more raw data from the test. Essentially everything is collected.) The TCP state collected by the NDT test is now specified in an IETF standard.[65]

In other words, this is an excellent testing tool and infrastructure. The insights to draw from this data, however, are not simple averages of the upload and download speeds from different user populations. This, in fact, would not be an appropriate use of the data as far too many factors confound such an interpretation. Rather the value of the NDT data is in understanding the sources of the performance bottlenecks for today's network users. Analyzing the publically available data[66] from this test has been very helpful in advancing our own understanding of the performance bottlenecks on today's broadband networks.

---

[60] FAQ available at http://www.google.com/support/youtube/bin/answer.py?answer=174122

[61] See http://fasterdata.es.net/expected.html

[62] See http://www.measurementlab.net/

[63] See http://blog.bittorrent.com/2010/01/25/easier-setup-in-%C2%B5torrent-2-0-and-measurement-lab-collaboration/

[64] See http://www.web100.org/

[65] RFC 4898, "TCP Extended Statistics MIB" http://www.ietf.org/rfc/rfc4898.txt

[66] http://developer.amazonwebservices.com/connect/entry.jspa?externalID=3190&categoryID=279

Understanding these local bottlenecks is indeed one of the prime reasons the test was built in the first place. Documentation on the NDT webpage[67] notes:

Several studies have shown that the majority of network performance problems occur in or near the users' desktop/laptop computer. These problems include, but are not limited to, duplex mismatch conditions on Ethernet […] *links, incorrectly set TCP buffers in the user's computer, or problems with the local network infrastructure. The NDT is designed to quickly and easily identify a specific set of conditions that are known to impact network performance.*

4.2.5.1. NDT Test Methodology

The primary NDT tests of interest for this paper are the inbound and outbound speed tests. (There are also checks for middleboxes and firewalls that are conducted.) The speed measured in both directions is the total bytes written/read from the server side socket divided by the test duration (therefore obviously a measure of a single TCP connection). This is an interesting slight measurement asymmetry for the tests in both directions as the bytes read by the server have successfully traversed the network. However the bytes written to the socket have not definitively made it to the client yet. So the download speed might be ever so slightly higher than the upload speed for simple methodological reasons. This effect is very negligible and we only point it out to demonstrate that even for the same testing tool "speed" may be measured in slightly different ways. The Ookla/Speedtest engine similarly measures both upload and download speeds from the same side for both tests so is also slightly asymmetric in its measurement approach. In both cases implementation reasons dictate how the speed is measured.

In the NDT test the sender tries to send as much data as possible during the ten second test period. This differs from the Akamai and Comscore approaches which time the download of a fixed amount of data but parallels the Ookla/Speedtest approach that will generally still be downloading data when the ten second test completes. As broadband speeds increase, tests that attempt to download as much data as possible have the potential to download a significant amount of traffic.

4.2.5.2. NDT Implementation

The NDT test consists of a client program that comes in the form of a Java applet and a command line version. The version running on the broadband.gov website has a more user friendly Flash based façade over the Java applet but the downside is that this hides most of the valuable explanatory data produced by the test. The entire source code for both the server and client side is publically available.[68] The server side consists of a webserver and testing and

[67] See http://www.internet2.edu/performance/ndt/

[68] See http://www.internet2.edu/performance/ndt/download.html

analysis engine. The key to the server side functionality is the instrumentation of the operating system's TCP stack which allows detailed connection statistics to be captured every 5 ms for every test. A packet trace of every test is also captured.

Clients can manually select the test server to connect or can be assigned automatically to a server based by the Donar (Distributed server selection for cloud services) infrastructure.[69] This automatic selection of servers is based upon both proximity of the client to different test servers and the load on the infrastructure as a whole. This is in contrast to the Ookla infrastructure where the automatic server selection is based only on geographic proximity but mirrors the Akamai infrastructure which similarly takes load into account.

4.2.5.3. NDT Public Data

All data from NDT tests run between February 2009 and September 2009 were made available by the M-labs project on the Amazon web services site in December of 2009.[70] This represents over 500 GB of test data. This is an impressive amount but dwarfed in size by the amount of NDT data that will be eventually publically available on a new Google service called BigQuery. As of May 2010, BigQuery contains M-Lab logs generated between February 2009 and May 2010 by the NDT test (and one other M-Lab test).[71] This consists of over 60 billion rows of data corresponding to 6 million broadband speed tests.[72]

To our knowledge there has not yet been an analysis of either of these data sets. We therefore analyzed all the data tests run between February 2009 and September 2009 in the first data set as that was all that was available. This consisted of approximately 380,000 tests of which 71,000 were in the US.[73] The median speed for all tests in the United States was 4.9 mbps and the average speed was 8.9 mbps. (See Figure 10 for a histogram of speeds.)

---

[69] See http://donardns.org

[70] See http://developer.amazonwebservices.com/connect/entry.jspa?externalID=3190&categoryID=279

[71] See http://code.google.com/apis/bigquery/docs/codelab-mlab.html

[72] See http://www.measurementlab.net/news/2010/apr/01/good-data-better-no-data-m-lab-built-continually-improve-data-collection-efforts

[73] There are approximately 400,000 tests in the actual data set. We excluded tests that for instance did not last between 10 and 11.5 seconds or were otherwise broken. The dataset of connections we analyzed is publically available to anyone interested in extending or verifying our analysis.

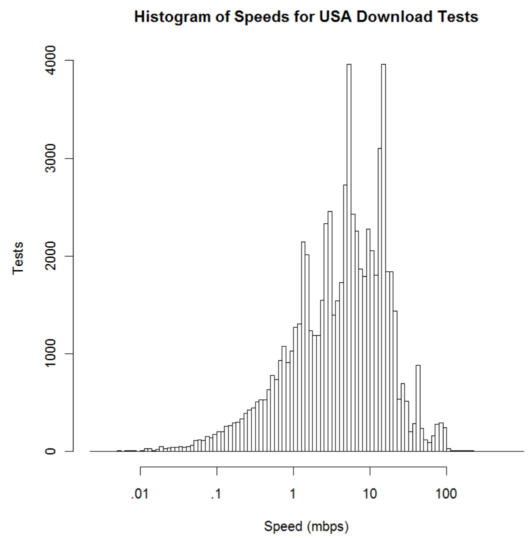

Figure 10: Histogram of speeds for download tests in the United States.

But as we stated above, these simple summary statistics tell us little about broadband networks as over one third of the tests (34% in the US and 38% in the entire dataset) never used all the available network capacity. The median and average speed for the uncongested tests in the United States was 5.5 mbps and 8.7 mbps respectively. For these tests, TCP never grew its sending rate to a level that caused any network congestion. More technically, the number of multiplicative downward congestion window adjustments due to all forms of congestion signals, including Fast Retransmit, ECN and timeouts was zero. The multiplicative decrease portion of TCP congestion control was never invoked.

What was constraining the sending rate if it wasn't TCPs response to network congestion for these tests? The answer is the receiver window (rwnd) advertised by TCP. The receiver window tells the sender how much buffer space the receiver has to store incoming data. This is generally determined by the operating system but can be tweaked by users or applications though this is not particularly common. If the receive window is the bottleneck there is a simple formula that predicts the speed that will be achieved -- the receiver window divided by the round trip time (Rwnd/RTT).

Figure 11 is a scatter plot of the measured download speeds for a test versus the average round trip time for a connection. The plot on the left is the subset of all tests that did not receive any congestion signals, on the right is all tests. The overlaid receiver window plot lines represent the theoretical speed predicted by the formula rwnd/average-RTT. Each test is color-coded by the range of the max receiver window TCP sent during the connection that the test falls into. For example, blue dots (near the third line from the top) correspond to tests with a max receiver window that is greater than 16KB and less than 64KB. (We actually shift each of these bucket boundaries upward by 10,000 bytes to account for differences in the windows advertised by different OSes.)

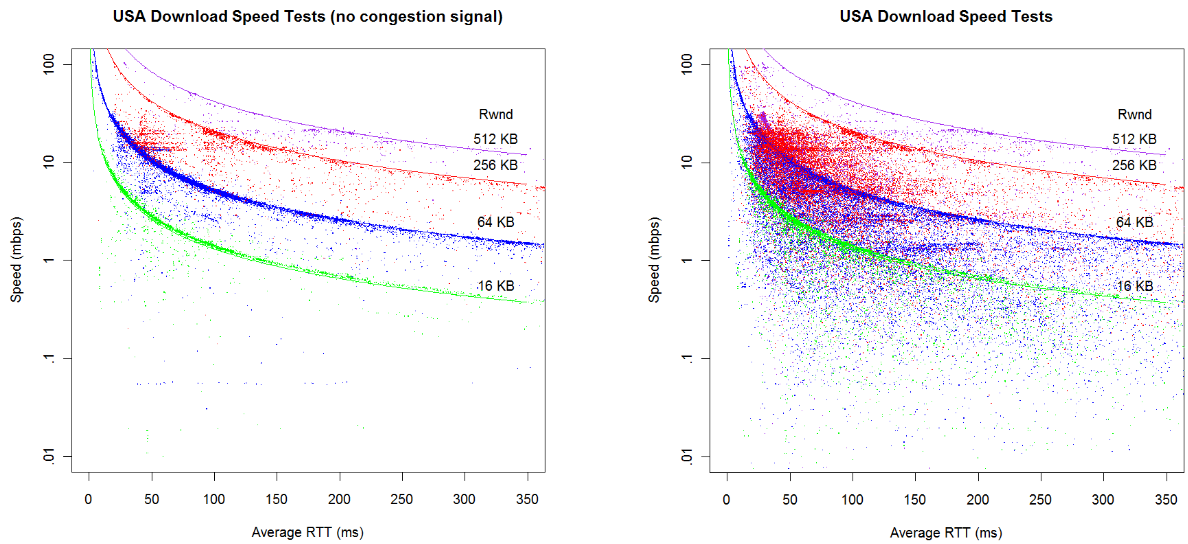

Figure 11: Scatter plots of the measured download speeds for a test versus the average round trip time for a connection. The plot on the left is the subset of all tests that did not receive any congestion signals, on the right is all tests. The receiver window (rwnd) plot lines represent the theoretical speed predicted by the formula rwnd/average-RTT. Each test is colored coded by the range the max receiver window TCP sent during the connection falls into. Blue dots (near the third line from the top) for instance have a max receiver window that is greater than 16KB and less than 64KB. (We actually shift each of these bucket boundaries upward by 10,000 bytes to account for slight difference in the rwnd advertised by different OSes.)

There are a number of lessons to take away from these plots.[74] The first is that the receiver window has a considerable effect on the speeds that can be achieved. If one looks at any vertical column of tests for clients that share similar round trip times, a main cause of differences is different receive windows. A second lesson is that the average round trip time has a considerable effect on the speed. Consider clients that have advertised similar receive windows (say, for instance, those in blue that have around a 64 KB window a common number for Windows XP boxes). The maximum speed measured for any of these drops off rapidly as the average round trip time increases. This demonstrates why server selection is so important for the speed that can be measured.

4.3. Applications and other websites which provide speed measurements

In this section we consider various applications and other websites which provide user visible speed measurements. These are measurements provided by applications like web browsers and websites that are likely found by users looking to test their speed. For example, the CNET

[74] There is a considerably number of other insights to be gathered from these NDT tests but we do not expand upon it here. We plan to include our expanded analysis in other papers.

bandwidth tester, which we consider below, is in the top ten Google search results obtained when one searches for "broadband speed test." While the ones we consider here don't feed into any aggregate statistics, they are often seen and noted by end-users.

When downloading a file in a web browser, the total transfer speed is often presented at the end of the connection like in Figure 12 below from the Internet Explorer which notes that the 5MB file was transferred in 3 seconds which is displayed as 1.66MB/sec or 13.3 mbps. What is interesting is that the packet capture shows that the transfer actually occurred within 2.2 seconds implying a throughput of 19.7 mbps. Evidently, Internet Explorer rounded up to the nearest second which for a short file transfer of 5MB represents a significant time increase impacting the overall measured speed by more than 6 mbps. (Incidentally, this transfer occurred on a 12mbps connection with 15mbps Powerboost.)

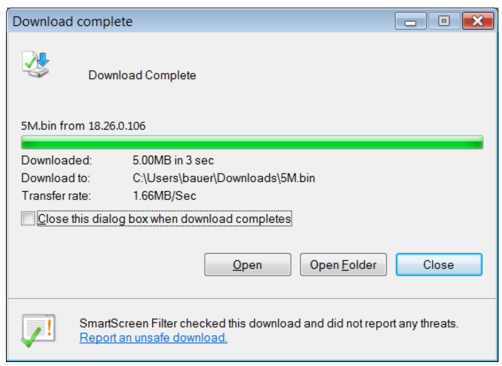

Figure 12: Internet Explorer dialog box that displays information that the file transfer rate was 13.3 mbps while the packet client side packet capture clearly demonstrates that the transfer rate was 19.7 mbps. The probable rounding up of the transfer time for this relatively small file was the likely source of this discrepency.

We find the CNet bandwidth meter to be wildly inaccurate. For a 12mbps with Powerboost to 15 mbps connection, CNet exclusively measures the download speeds as being below 2 mbps (see Figure 13). They calculate this speed by timing the download of a single image file using Javascript. While there are other sources of inaccuracies[75] the most significant in our view is that the file size is far too small particularly given the inaccuracies of timing the download in Javascript.

[75] See http://www.rankingtoday.com/net/speed-test-comparison.php for a detailed examination of CNet and other browser based speed testing sites.

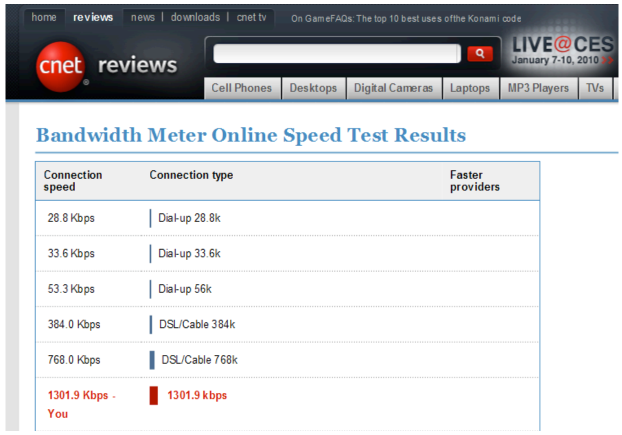

Figure 13: CNet bandwidth test conducted on a 12mbps cable modem.

These sources of misinformation make it potentially confusing for end users to understand how the broadband service they buy is actually performing. The broadband access networks certainly can be the source of performance problems and bottlenecks, but there are many other sources of problems as well.

5. Why speed matters

While we hopefully have convinced the reader from the preceding discussion that there is no definitive definition of a "right" measure for broadband speed, it is worth emphasizing that speed will remain an important and focal statistic for characterizing broadband service. Speed *does* matter but it is not all that matters.

There are a number of reasons for why so much analysis and discussion focuses on speed. Broadband speed is one of the most significant technical characteristics impacting the quality of the user experience (how well different applications may run). The raw data rate that broadband supports is important for the design of applications (e.g,. video encoding rates), devices (e.g., I/O speeds supported), and complementary infrastructure (e.g,. transport backhaul, required buffer memory, and switching fabric design). According to one report: "While 93% of broadband users claim to be satisfied with their experience of web browsing, satisfaction rates are lower for all other services which require the streaming or downloading of content and therefore benefit from higher speeds and/or more consistent performance."[76] Thus, a lack of availability of sufficiently high-speed broadband access services may be responsible for holding back the next generation of Internet applications and innovation.

Broadband speeds provide a rough way to characterize the type of infrastructure and may be used as shorthand for the level of investment needed or the quality of existing services. Although

---

[76] http://www.ofcom.org.uk/research/telecoms/reports/bbspeed_jan09/bbspeed_jan09.pdf

speed is not the only determinant of the technical quality of service, it is one of the important characteristics and is often positively correlated with other indices of service quality (e.g., latency, jitter, packet loss, etcetera).[77]

The market has focused on advertised "up to" peak download speeds as the single most often-cited technical characteristic for differentiating service offerings. Service providers typically tier their service offerings based on price and speed (and other service terms that may be described in the small print) with services offering higher peak data rates being offered at higher prices. Customers self-select into appropriate speed tiers based on budget considerations (how much can they afford to spend for broadband?) and expected cost-benefit tradeoffs (how much better performance will be realized and how much will that matter if a higher speed/more expensive service is selected?). Consumers with different usage profiles may compute the cost-benefit tradeoffs differently (e.g., heavy video users may value speed more than light email users). Moreover, consumers differ widely in sophistication about the technical characteristics of different services and their willingness-to-pay for broadband service evolves over time.

Policymakers have also focused on speed as a metric of merit for defining/characterizing broadband services. It is used to segment broadband service markets. This renders the appropriate characterizations of speed relevant for market power assessments and potentially for defining quality-of-service standards and identifying what services qualify for what type of regulatory treatment.[78] The FCC's broadband data collection efforts currently differentiate broadband according to speed tier combinations of upload and download speeds.[79]

In the following sub-sections, we consider some of the ways in which policymakers may make use of speed measurements and data.

5.1. What can be advertised?

One question policymakers face is whether a speed that is advertised has to be a speed that can be actually measured with a test. Personally, we would not like to see a 100 mbps Ethernet link have to be advertised as a 96 mbps speed tier (because one can never measure a 100 mbps TCP

[77] Latency, jitter, and packet loss are all phenomena that occur because queues of packets build up within network devices. For a fixed level of demand, a faster link will be capable of draining queues more quickly (or preventing a build up in the first place) thus reducing the latency, jitter, and the potential for queues to overflow resulting in packet loss.

[78] The new European regulatory framework that was adopted in the fall of 2009 allows national regulatory authorities to define minimum quality of service standards for broadband. If adopted, any such definitions will surely include reference to broadband speeds. More generally, if broadband-specific regulations are adopted, there will be a need to determine what services qualify as "broadband" service and what do not.

[79] See "FCC Releases new Census tract-level data on high-speed Internet services," Press Release, Federal Communications Commission, February 12, 2010, available at: http://www.fcc.gov/Daily_Releases/Daily_Business/2010/db0212/DOC-296234A1.pdf. The FCC's data collection is based on advertised transmission speeds and includes services offering at least 200 kbps transmission speeds.

throughput on such a link). So a government rule that said one can only advertise a given speed if that speed can be measured with a particular testing methodology would be undesirable from our perspective. The advertisement of access link speeds or provisioned speeds (or something below either) should be acceptable. Noting that a service is 100 mbps is useful to technically sophisticated customers and consumer advocates who would generally understand that to be a 100 mbps Ethernet.

However, the example from the previous paragraph becomes more complex if scaled down to lower speed service offerings. If a provider advertised up to 10 mbps but only delivered 5 mbps most of the time (measured with a test that carefully identified the network performance of the broadband provider) we would be sympathetic to a consumer that this was potentially misleading. Particularly for some broadband technologies, like wireless and DSL, it can be very difficult for a provider or consumer to tell a priori what speeds will be possible. In the case of DSL, the distance to the upstream equipment has a very significant effect of the speeds that can be achieved. This is frustrating, but partially unavoidable, for consumers that have to try the service to understand what they will actually experience. For networks that are not as sensitive to the distance from the subscriber to the upstream aggregation point, a large consistent discrepancy between the advertised speed and achieved speed is more troubling.

5.2. Broadband benchmarking

It is reasonable to anticipate that speed measurements averaged over geographic areas or provider networks may be used for benchmarking. Such benchmarking, if sufficiently accurate, might be useful for targeting broadband stimulus funding or to evaluate earlier investments. It could be useful to help identify service problems or to highlight best practices.

Because of the strategic potential for such data to influence behavior (e.g., induce consumers to switch to another broadband service provider or policymakers to target remedial actions), there is a concern that the data be (a) accurate and (b) immune from manipulation. For example, if two ISPs are compared, it is important that the measurements actually focus on the services offered by the two ISPs rather than other exogenous factors that impact performance (e.g., potential differences in the user behavior of their customer base which may be reflected in systematic differences in the applications used, destination servers, or configuration of user-premise equipment). Furthermore, any comparisons need to be contextually relevant. For example, suppose daily rankings show two ISPs flip-flopping, then while it might look as if one ISP was providing better service on any particular day, there would not be any evidentiary evidence for identifying a better ISP over time. The appropriate presentation and interpretation of data will depend, in part, on what the data actually shows (not just statistical means, but also variances and higher statistical moments).

We regard as highly desirable efforts to implement special test infrastructures and tools such as those offered by SamKnows, the consultancy that OfCom employed for its assessment of the condition of broadband services in the United Kingdom. The SamKnows approach involves deploying a set of measurement servers and testing nodes that allows for tighter isolation of broadband access connection behavior than is afforded by other approaches. While this approach offers better isolation and control over many aspects of what one might like to measure, it is not without its own limitations. The population of potential test sites is much more limited. So while we believe the SamKnows approach is a useful complement to other measurements, we expect other measurements to continue to be useful as well. Multiple testing methods and infrastructure will provide unique insights from each as well as enable cross-validation of results.

6. Conclusion

Broadband access networks are the vital bridges tying users to the communication, content, entertainment, and marketplaces of the Internet. A healthy broadband service market will include a differentiated menu of service offerings from (hopefully) multiple broadband ISPs. For efficiency, market participants will need information about the range of choices available, the associated pricing, and information about the quality of service associated with the different offerings. Today, speed or data rate is the single most important technical metric for characterizing broadband service with faster speed equating to better performance (holding other characteristics like price constant).

However, in the next few years, as the average speed of broadband increases, and the markets become more sophisticated, we expect that attention may shift towards a more nuanced characterization of what matters for evaluating the quality of broadband services. Issues such as availability (reliability) and latencies to popular content and services may become more important in how services are advertised and measured. We welcome such a more nuanced view and believe it is important even in so far as one's principal focus is on broadband speeds.

The main objective of this paper has been to facilitate the informed discussion of broadband performance and advocate for a more nuanced interpretation of the speed data that is cited today. We have highlighted different definitions for broadband speed. This variability in what one wants from a broadband speed measure helps account for some of the variability in speed metrics. There is no single best measurement approach for all situations.

We present a detailed analysis of some of the sources of methodological differences associated with several of the more popular measurement sources in use today. In addition to its value in enhancing our understanding of public data sources, these methodological differences provide a cautionary tale for those who are too quick to quote empirical measurement results without understanding precisely how those measurements are derived.

We hope that this paper and the associated research will contribute to an active and better informed discussion among all industry participants about speed metrics. We believe competitive broadband markets will work better with more information and better data -- and *better understood* data -- about broadband speeds.