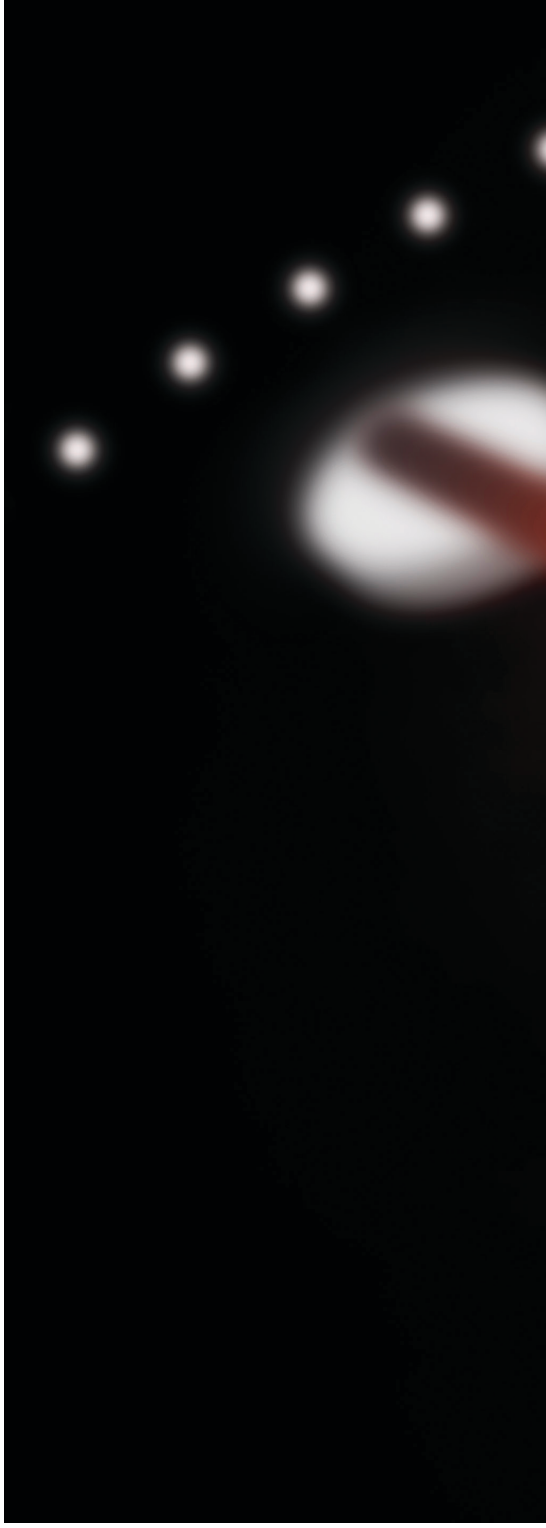
DOI:10.1145/3372135

Speed testing methods have flourished over the last decade, but none without at least some limitations.

BY NICK FEAMSTER AND JASON LIVINGOOD

Measuring Internet Speed

Current Challenges and Future Recommendations



VARIOUS GOVERNMENTAL ORGANIZATIONS have begun to rely on so-called Internet speed tests to measure broadband Internet speed. Examples of these programs include the Federal Communications Commission's "Measuring Broadband America" program,[7] California's CALSPEED program,[4] the U.K.'s Home Broadband Performance Program,[23] and various other initiatives in states including Minnesota,[18] New York,[19–21] and Pennsylvania.[27] These programs have various goals, ranging from assessing whether ISPs are delivering on advertised speeds to assessing potentially underserved rural areas that could benefit from broadband infrastructure investments.

The accuracy of measurement is critical to these assessments, as measurements can inform everything from investment decisions to policy actions and even litigation. Unfortunately, these efforts sometimes rely on outmoded technology, making the resulting data unreliable or misleading. This article describes the current state of speed testing tools, outlines their limitations, and explores paths forward to better inform the various technical and policy ambitions and outcomes.

Some current speed test tools were well-suited to measuring access link capacity a decade ago but are no longer useful because they made a design assumption that the Internet Service Provider (ISP) last mile access net-

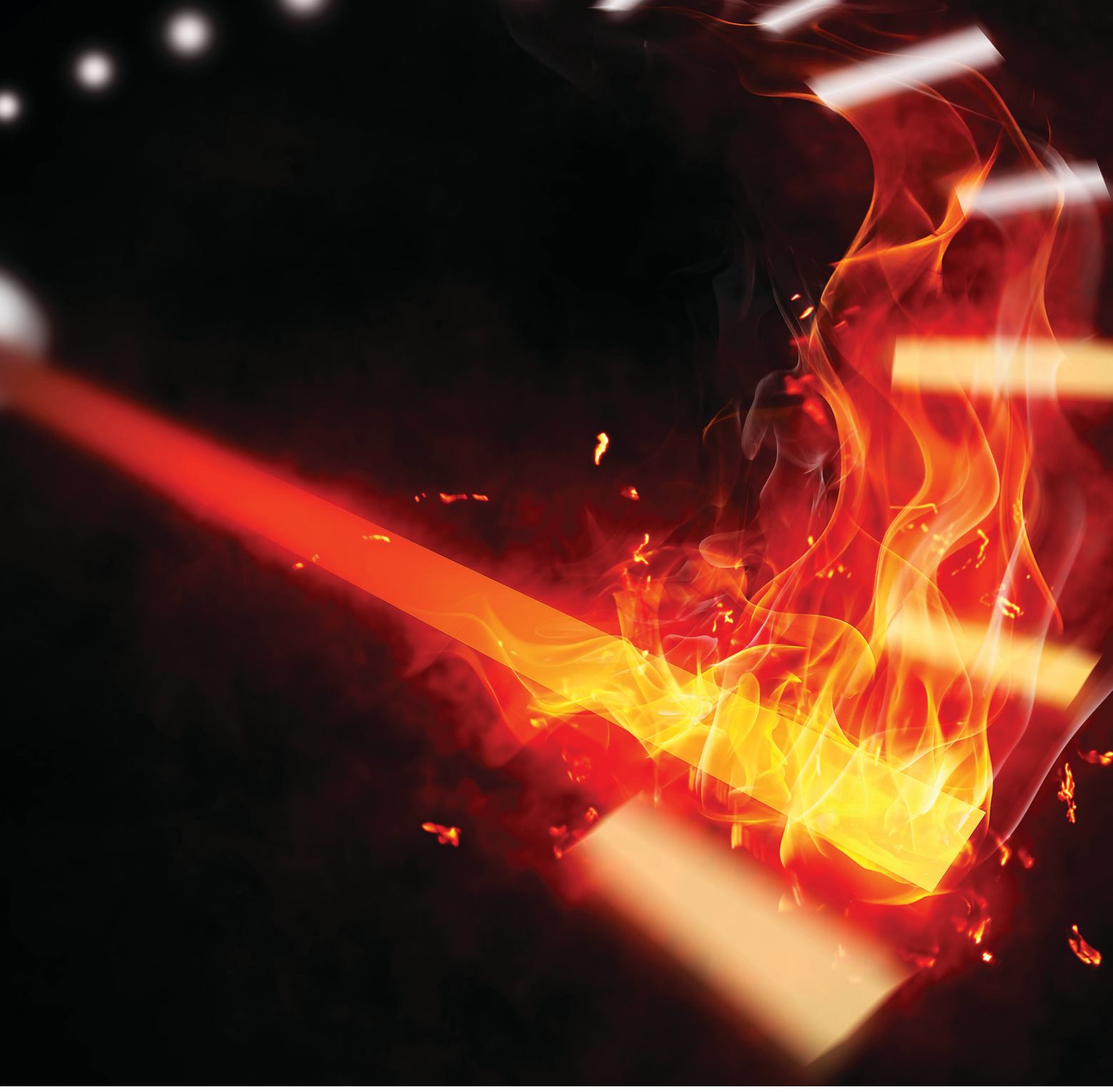


IMAGE BY YUGANOV KONSTANTIN

work was the most constrained (bottleneck) link. This is no longer a good assumption, due to the significant increases in Internet access speeds due to new technologies. Ten years ago, a typical ISP in the United States may have delivered tens of megabits per second (Mbps). Today, it is common to have ten times faster (hundreds of megabits per second), and gigabit speeds are available to tens of millions of homes. The performance bottleneck has often shifted from the ISP access network to a user's device, home Wi-Fi network, network interconnections, speed testing infrastructure, and other areas.

A wide range of factors can influence the results of an Internet speed test, including: user-related considerations, such as the age of the device; wide-area network considerations, such as interconnect capacity; test-infrastructure considerations, such as test server capacity; and test design, such as whether the test runs while the user's access link is otherwise in use. Additionally, the typical Web browser opens multiple connections in parallel between an end user and the server to increasingly localized content delivery networks (CDNs), reflecting an evolution of applications that ultimately effects the user experience.

These developments suggest the need to evolve our understanding of the utility of existing Internet speed test tools and consider how these tools may need to be redesigned to present a more representative measure of a user's Internet experience.

Background

In this section, we discuss and define key network performance metrics, introduce the general principles of Internet "speed tests" and explore the basic challenges facing any speed test.

Performance metrics. When people talk about Internet "speed," they are generally talking about throughput. End-to-end Internet performance is typically measured with a collection of metrics—specifically throughput (that is, "speed"), latency, and packet loss. Figure 1 shows an example speed test from a mobile phone on a home Wi-Fi network. It shows the results of a "native" speed test from the Ookla Android speed test application[24] run in New Jersey, a canonical Internet speed test. This native application reports the user's ISP, the location of the test server destination, and the following performance metrics:

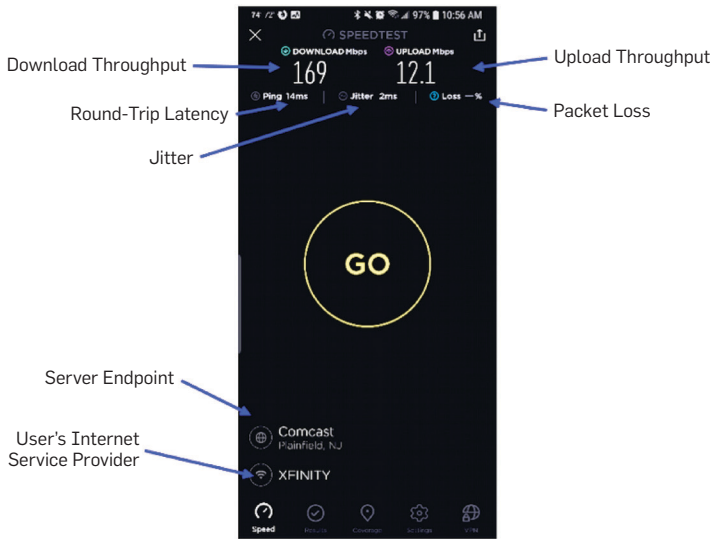
Throughput is the amount of data that can be transferred between two network endpoints over a given time interval. For example, throughput can be measured between two points in a given ISP's network, or it can be measured for an end-to-end path, such as between a client device and a server at some other place on the Internet. Typically, a speed test measures both downstream (download), from server to client, and upstream (upload), from client to server (Bauer et al.[2]) offer an in-depth discussion of throughput metrics). Throughput is not a constant; it changes from minute to minute based on many factors, including what other users are doing on the Internet. Many network-performance tests, such as the FCC test[7] and Ookla's speed test, include additional metrics that reflect the user's quality of experience.

Latency is the time it takes for a single data packet to travel to a destination. Typically, latency is measured in terms of *roundtrip latency*, since measuring one-way latency would require tight time synchronization and the ability to instrument both sides of the Internet path. Latency generally increases with distance, due to factors such as the speed of light for optical network segments; other factors can influence latency, including the amount of queueing or buffering along an end-to-end path, as well as the actual network path that traffic takes from one endpoint to another. TCP throughput is inversely proportional to end-to-end latency;[31] all things being equal, then, a client will see a higher throughput to a nearby server than it will to a distant one.

Jitter is the variation between two latency measurements. Large jitter measurements are problematic.

Packet loss rate is typically computed as the number of lost packets divided by the number of packets transmitted. Although high packet loss rates generally correspond to worse performance, some amount of packet loss is normal because a TCP sender typically uses packet loss as the feedback signal to determine the best transmission rate. Many applications such as video streaming are designed to adapt well to packet loss without noticeably affecting the end user experience, so there is no single level of packet loss that automatically translates to poor application performance. Additionally, certain network design choices, such as increasing buffer sizes, can reduce packet loss, but at the expense of latency, leading to a condition known as "buffer bloat."[3,12]

Figure 1. Example metrics from an Ookla Speedtest, a canonical Internet speed test.



Speed test principles and best practices. *Active measurement.* Today's speed tests are generally referred to as active measurement tests, meaning that they attempt to measure network performance by introducing new traffic into the network (so-called "probe traffic"). This is in contrast to *passive* tests, which observe traffic passing over a network interface to infer performance metrics. For speed testing, active measurement is the recognized best practice, but passive measurement can be used to gauge other performance factors, such as latency, packet loss, video quality, and so on.

Measuring the bottleneck link. A typical speed test sends traffic that traverses many network links, including the Wi-Fi link inside the user's home network, the link from the ISP device in the home to the ISP network, and the many network level hops between the ISP and the speed test server, which is often hosted on a network other than the access ISP. The throughput measurement that results from such a test in fact reflects the capacity of the *most constrained* link, sometimes referred to as the "bottleneck" link—the link along the end-to-end path that is the limiting factor in end-to-end throughput. If a user has a 1Gbps connection to the Internet but their home Wi-Fi network is limited to 200Mbps, then any speed test from a device on the Wi-Fi network to the Internet will not exceed 200Mbps. Bottlenecks can exist in an ISP access network, in a transit network between a client and server, in the server or server data-center network, or other places. In many cases, the bottleneck is located somewhere along the end-to-end path that is not under the ISP's or user's direct control.

Use of transmission control protocol. Speed tests typically use the Transmis-

Figure 2. TCP Dynamics.

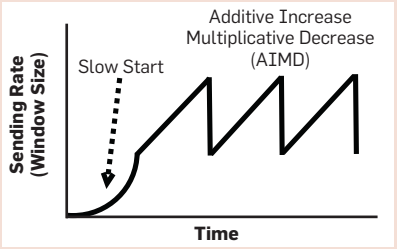
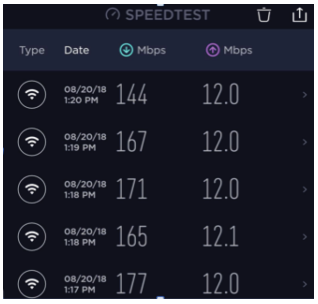
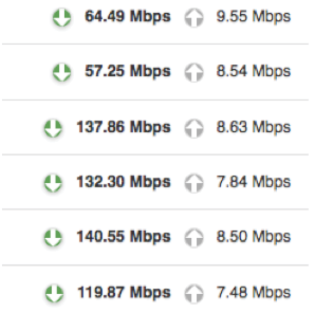


Figure 3. Successive runs of different throughput tests.



(a) Five successive runs of Ookla Speedtest yield variable results on downstream throughput.



(b) Internet Health Test runs in succession to six different servers. The test measures consistently lower throughput and also shows variability, both to different servers and across successive test runs.

sion Control Protocol (TCP) to measure throughput. In keeping with the nature of most Internet application transfers today—including, most notably, Web browsers—most speed tests use multiple parallel TCP connections. Understanding TCP's operation is critical to the design of an accurate speed test. Any TCP-based speed test should be: long enough to measure steady-state transfer; recognize that TCP transmission rates naturally vary over time; and, use multiple TCP connections. Figure 2 shows TCP's dynamics, including the initial slow start phase. During TCP slow start, the transmission rate is far lower than the network capacity. Including this period as part of a throughput calculation will result in a throughput measurement that is less than the actual available network capacity. If test duration is too short, the test will tend to underestimate throughput. As a result, accurate speed test tools must account for TCP slow start. Additionally, instantaneous TCP throughput continually varies because the sender tries to increase its transfer rate in an attempt to find and use any spare capacity (a process known as "additive increase multiplicative decrease" or AIMD).

Inherent variability. A speed test measurement can produce highly variable results. Figure 3 shows an illustrative example of typical variability that a speed test might yield, both for Internet Health Test (IHT) and Ookla Speedtest. These measurements were performed successively on the same Comcast connection provisioned for 200Mbps downstream and 10Mbps upstream throughput. The tests were performed in succession. Notably, successive tests yield different measurements. IHT, a Web front-end to a tool called the Network Diagnostic Test (NDT), also consistently and significantly under-reports throughput, especially at higher speeds.

Limitations of Existing Speed Tests

Existing speed tests have a number of limitations that have become more acute in recent years, largely as a result of faster ISP access links and the proliferation of home wireless networks. The most profound change is that *as network access links have become faster, the network bottleneck has moved from the ISP access link to elsewhere on the network*. A decade ago, the network bottleneck was commonly the access ISP link; with faster ISP access links, the network bottleneck may have moved any number of places, from the home wireless network to the user's device itself. Other design factors may also play a role, including how measurement samples are taken and the provisioning of the test infrastructure itself.

User-related consideration. *The home wireless network.* Speed tests that are run over a home wireless connection often reflect a measurement of the user's home wireless connection, *not* that of the access ISP, because the Wi-Fi network itself is usually the lowest capacity link between the user and test server.[1,5,16,26,28,30] Many factors affect the performance of the user's home wireless network, including: distance to the Wi-Fi Access Point (AP) and Wi-Fi signal strength, technical limitation of a wireless device and/or AP, other users and devices operating on the same network, interference from nearby APs using the same spectrum, and interference from non-Wi-Fi household devices that operate on the same spectrum (for example, microwave ovens, baby monitors, security cameras).

Many past experiments demonstrate that the user's Wi-Fi—not the ISP—is often the network performance bottleneck. Sundaresan et al. found that whenever downstream throughput exceeded 25Mbps, the user's home wireless network was almost always the bottleneck.[30] Although the study is from 2013, and both access link speeds and wireless network speeds have since increased, the general trend of home wireless bottlenecks is still prevalent.

Client hardware and software. Client types range from dedicated hardware, to software embedded in a device on the user's network, to native software made for a particular user operating system, and Web browsers. Client type has an important influence on the test results, because some may be inherently limited or confounded by user factors. Dedicated hardware examples include the SamKnows whitebox and RIPE Atlas probe. Embedded software refers to examples where the software is integrated into an existing network device such as cable modem, home gateway device, or Wi-Fi access point. A native application is software made specifically to run on a given operating system such as Android, iOS, Windows, and Mac OS. Finally, Web-based tests simply run from a Web browser. In general, dedicated hardware and embedded software approaches tend to be able to minimize the effect of user-related factors and are more accurate as a result.

Many users continue to use older wireless devices in their homes (for example, old iPads and home routers) that do not support higher speeds. Factors such as memory, CPU, operating system, and network interface card (NIC) can significantly affect throughput measurements. For example, if a user has a 100Mbps Ethernet card in their PC connected to a 1Gbps Internet connection, their speed tests will never exceed 100Mbps and that test result cannot be said to represent a capacity issue in the ISP network; it is a device limitation. As a result, many ISPs document recommended hardware and software standards,[33] especially for 1Gbps connections. The limitations of client hardware can be more subtle. Figure 4 shows an example using iPhone released in 2012–2015. This shows that any user with an iPhone 5s or older is unlikely to reach 100Mbps, likely due to the lack of a newer 802.11ac wireless interface.

Figure 4. Distribution of download speeds across different device types. Older devices do not support 802.11ac, so fail to consistently hit 100Mbps.

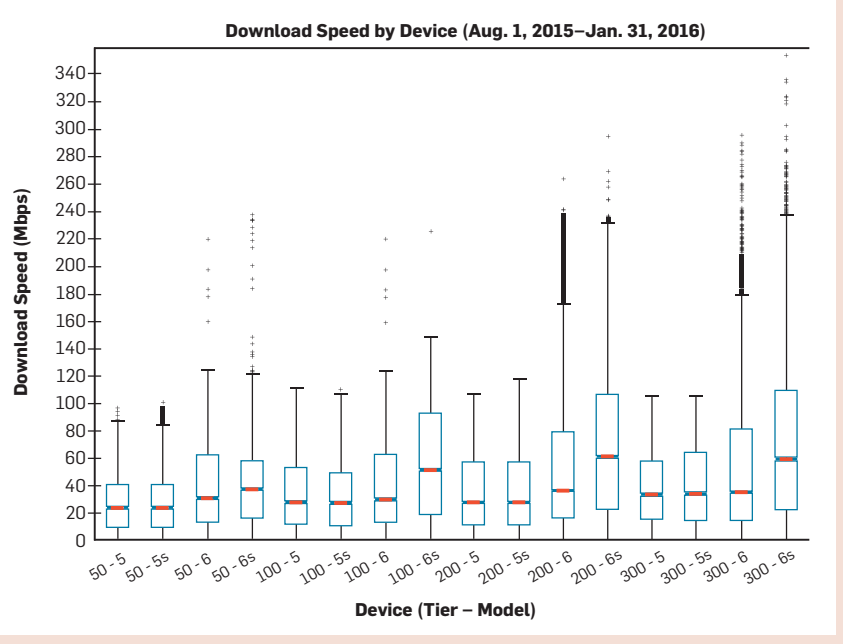
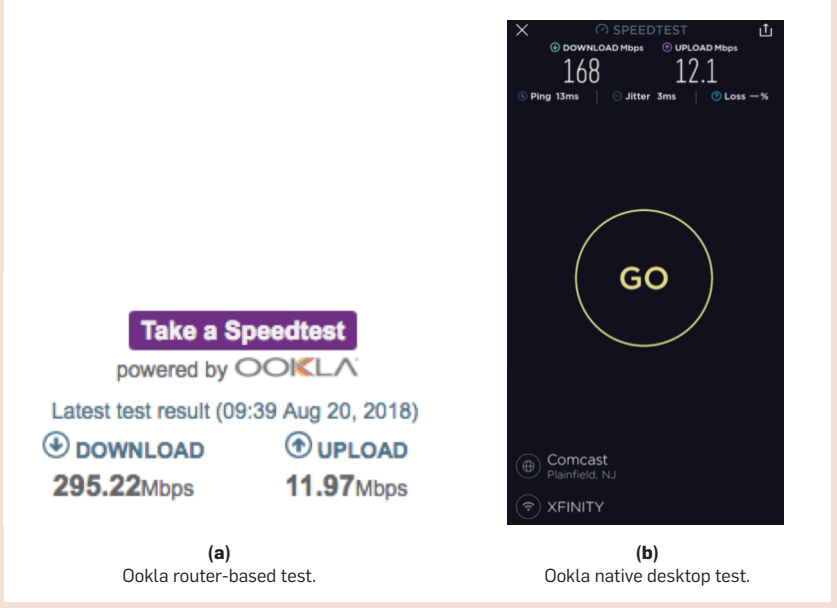


Figure 5. Two forms of Ookla Speedtest: A router-based test and native desktop test from the same home network.



(a) Ookla router-based test.

(b) Ookla native desktop test.

Router-based testing vs. device-based testing. Figure 5 shows an example of two successive speed tests. Figure 5a uses software embedded in the user's router, so that no other effects of the local network could interfere. Figure 5b shows the same speed test (such as, Ookla Speedtest), on the same network, performed immediately following the router-based test using native software on a mobile device over Wi-Fi. The throughput reported from the user's mobile device on the home network is almost half of the throughput that is reported when the speed test is taken directly from the router.

Competing "cross traffic." At any given time, a single network link is simultaneously carrying traffic from many senders and receivers. Thus, any single network transfer must share the available capacity with the competing traffic from other senders—so-called *cross traffic*. Although sharing capacity is natural for normal application traffic, a speed test that shares the available capacity with competing cross traffic will naturally underestimate the total available network capacity. Client-based speed tests cannot account for cross traffic; because the client cannot see the volume of other traffic on the same network, whereas a test that runs on the user's home router can account for cross traffic when conducting throughput measurements.

Wide-area network considerations. *Impaired ISP access network links.* An ISP's "last mile" access network links can become impaired. For example, the quality of a DOCSIS connection to a home can become impaired by factors such as a squirrel chewing through a line or a bad ground wire. Similarly, fixed wireless connections can be impaired by weather or leaves blocking the antenna. To mitigate the potential for an individual impairment unduly influencing ISP-wide results, tests should be conducted with a large number of users.

Access ISP capacity. Capacity constraints within an ISP's network can exist, whether in the access network, regional network (metropolitan area), or backbone network. Regional and back-

bone networks usually have excess capacity so the only periods when they may be constrained would be the result of a disaster (for example, hurricane damage) or temporary conditions such fiber cuts or BGP hijacking. Usually ISP capacity constraints arise in the last-mile access networks, which are by nature shared in the first mile or first network element, (for example, passive optical networking (PON), DOCSIS, DSL, 4G/5G, Wi-Fi, point-to-point wireless).

Transit and interconnect capacity. Another significant consideration is the connection to "transit" and "middle mile" networks. The interconnects between independently operated networks may also introduce throughput bottlenecks. As user speeds reach 1Gbps, ensuring that there are no capacity constraints on the path between the user and test server— especially across transit networks—is a major consideration. In one incident in 2013, a bottleneck in the Cogent transit network reduced NDT throughput measurements by as much as 90%. Test results improved when Cogent began prioritizing NDT test traffic over other traffic. Transit-related issues have often affected speed tests. In the case of the FCC's MBA platform, this prompted them to add servers on the Level 3 network to isolate the issues experienced with M-Lab's infrastructure and the Cogent network, and M-Labs has also added additional transit networks to reduce their reliance on one network.

Middleboxes. End-to-end paths often have devices along the path, called "middleboxes," which can affect performance. For example, a middlebox may perform load balancing or security functions (for example, malware detection, firewalls). As access speeds increase, the capacity of middleboxes may increasingly be a constraint, which will mean test results will reflect the capacity of those middleboxes rather than the access link or other measurement target.

Rate-limiting. Application-layer or destination-based rate limiting, often referred to as throttling, can also cause the performance that users experience to diverge from conventional speed tests. Choffnes et al. have developed Wehe, which detects application-layer rate limiting;[32] thus far, the research has focused on HTTP-based video streaming de-prioritization and rate-limiting. Such rate limiting could exist at any point on the network path, though most commonly it may be expected in an access network or on the destination server network. In the latter case, virtual servers or other hosted services may be priced by peak bitrate and therefore a hard-set limit on total peak bitrate or per-user-flow bitrate may exist. Web software such as Nginx has features for configuring rate limiting,[22] as cloud-based services may charge by total network usage or peak usage; for example, Oracle charges for total bandwidth usage,[25] and FTP services often enforce per-user and per-flow rate limits.[11]

Many past experiments demonstrate that the user's Wi-Fi—not the ISP—is often the network performance bottleneck.

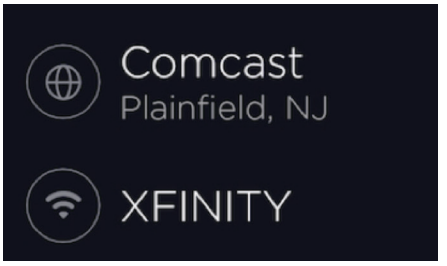
Rate-boosting. Rate-boosting is the opposite of rate limiting; it can enable a user to temporarily exceed their normal provisioned rate for a limited period. For example, a user may have a 100Mbps plan but may be allowed to burst to 250Mbps for limited periods if spare capacity exists. This effect was noted in the FCC's first MBA report in 2011 and led to use of a longer duration test to measure "sustained" speeds.[8] Such rate-boosting techniques appear to have fallen out of favor, perhaps partly due greater access speeds or the introduction of new technologies such as DOCSIS channel bonding.

Test infrastructure considerations. Because speed tests based on active measurements rely on performing measurements to some Internet endpoint (that is, a measurement server), another possible source of a performance bottleneck is the server infrastructure itself.

Test infrastructure provisioning. The test server infrastructure must be adequately provisioned so that it does not become the bottleneck for the speed tests. In the past, test servers have been overloaded, misconfigured, or otherwise not performing as necessary, as has been the case periodically with M-Lab servers used for both FCC MBA testing and NDT measurements. Similarly, the data center switches or other network equipment to which the servers connect may be experiencing technical problems or be subject to other performance limitations. In the case of the FCC MBA reports, at one point this resulted in discarding of data collected

Figure 6. IHT and Ookla geolocation.

(a)
Internet Health Test mistakenly locating a client in Princeton, N.J., to Philadelphia, P.A. (50+ miles away), and performing a speed test to a server in New York City.



(b)
Ookla Speedtest directing a client in Princeton, N.J., to an on-net Speedtest server in Plainfield, N.J. Ookla also allows a user to select another nearby server.

from M-Lab servers due to severe impairments.[6,9] The connection between a given datacenter and the Internet may also be constrained, congested, or otherwise technically impaired, as was the case when some M-Lab servers were single-homed to a congested Cogent network. Finally, the servers themselves may be limited in their capacity: if, for example, a server has a 1Gbps Ethernet connection (with real-world throughput below 1Gbps) then the server cannot be expected to measure several simultaneous 1Gnps or 2Gbps tests. Many other infrastructure-related factors can affect a speed test, including server storage input and output limits, available memory and CPU, and so on. Designing and operating a high scale, reliable, high-performance measurement platform is a difficult task, and as more consumers adopt 1Gbps services this may become even more challenging.[17]

Different speed test infrastructures have different means for incorporating measurement servers into their infrastructure. Ookla allows volunteers to run servers on their own and contribute these servers to the list of possible servers that users can perform tests against. Ookla uses empirical measurements over time to track the performance of individual servers. Those that perform poorly over time are removed from the set of candidate servers that a client can use. Measurement Lab, on the other hand, uses a fixed, dedicated set of servers as part of a closed system and infrastructure. For many years, these servers have been: constrained by a 1Gbps uplink; shared with other measurement experiments (recently, Measurement Lab has begun to upgrade to 10Gbps uplinks). Both of these factors can and did contribute to the platform introducing its own set of performance bottlenecks.

Server placement and selection. A speed test estimates the available capacity of the network between the client and the server. Therefore, the throughput of the test will naturally depend on the distance between these endpoints as measured by a packet's round trip time (RTT). This is extremely important, because TCP throughput is inversely proportional to the RTT between the two endpoints. For this reason, speed test clients commonly attempt to find the "closest" throughput measurement server to provide the most accurate test result and why many speed tests such as Ookla's, use thousands of servers distributed around the world. to select the closest server, some tests use a process called "IP geolocation," whereby a client location is determined from its IP address. Unfortunately, IP geolocation databases are notoriously inaccurate, and client location can often be off by thousands of miles. Additionally, latency resulting from network distance typically exceeds geographic distance, since network paths between two endpoints can be circuitous, and other factors such as network congestion on a path can affect latency. Some speed tests mitigate these effects with additional techniques. For example, Ookla's Speedtest uses IP geolocation to select an initial set of servers that are likely to be close, and then the client selects from that list the one with the lowest RTT (other factors may also play into selection, such as server network capacity). Unfortunately, Internet Health Test (which uses NDT) and others rely strictly on IP geolocation.

Figure 6 shows stark differences in server selection between two tests: Internet Health Test (which relies on IP geolocation and has a smaller selection of servers); and Ookla Speedtest (which uses a combination of IP geolocation, GPS-based location from mobile devices, and RTT-based server selection to a much larger selection of servers). Notably, the Internet Health Test not only mis-locates the client (determining that a client in Princeton, New Jersey is in Philadelphia), but it also selects a server that is in New York City, which is more than 50 miles from Princeton. In contrast, the Ookla test, which selects an on-network Comcast server in Plainfield, NJ, which is merely 21 miles away, and also gives the user the option of using closer servers through the "Change Server" option.

Test design cConsiderations. *Number of parallel connections.* A significant consideration in the design of a speed test is the number of parallel TCP connections that the test uses to transfer data between the client and server, since the goal of a speed test is to send as much data as possible and this is usually only possible with multiple TCP connections. Using multiple connections in parallel allows a TCP sender to more quickly and more reliably achieve the available link capacity. In addition to achieving a higher share of the available capacity (because the throughput test is effectively sharing the link with itself), a transfer using multiple connections is more resistant to network disruptions that may result in the sender re-entering TCP slow start after a timeout due to lost packets.

A single TCP connection cannot

typically achieve a throughput approaching full link capacity, for two reasons: a single connection takes longer to send at higher rates because TCP slow start takes longer to reach link capacity, and a single connection is more susceptible to temporarily slowing down transmission rates when it experiences packet loss (a common occurrence on an Internet path). Past research concluded that a speed test should have at least four parallel connections to accurately measure throughput.[29] For the same reason, modern Web browsers typically open as many as six parallel connections to a single server in order to maximize use of available network capacity between the client and Web server.

Test duration. The length of a test and the amount of data transferred also significantly affect test results. As described previously, a TCP sender does not immediately begin sending traffic at full capacity but instead begins in TCP slow start until the sending rate reaches a pre-configured threshold value, at which point it begins AIMD congestion avoidance. As a result, if a transfer is too short, a TCP sender will spend a significant fraction of the total transfer in TCP slow start, ensuring the transfer rate will fall far short of available capacity. As access speeds increase, most test tools have also needed to increase test duration.

Throughput calculation. The method that tests use to calculate results appears to vary widely; often this method is not disclosed. Tests may discard some high and/or low results, may use the median or the mean, may take only the highest result and discard the rest, and so on. This makes different tests difficult to compare. Finally, some tests may include all of the many phases of a TCP transfer, even though some of those phases are necessarily at rates below the capacity of a link:

- the slow start phase at the beginning of a transfer (which occurs in every TCP connection);
- the initial "additive increase" phase of the TCP transfer when the sender is actively increasing its sending rate but before it experiences the first packet loss that results in multiplicative decrease;
- any packet loss episode which results in a TCP timeout, and subsequent re-entry into slow start

Estimating the throughput of the link is not as simple as dividing the amount of data transferred by the total time elapsed over the course of the transfer. A more accurate estimate of the transfer rate would instead measure the transfer during steady-state AIMD, excluding the initial slow start period. Many standard throughput tests, including the FCC/SamKnows test, omit the initial slow start period. The Ookla test implicitly omits this period by discarding low-throughput samples from its average measurement. Tests that include this period will result in a lower value of average throughput than the link capacity can support in steady state.

Self-selection bias. Speed tests that are initiated by a user suffer from self-selection bias:[14] many users initiate such tests only when they are experiencing a technical problem or are reconfiguring their network. For example, when configuring a home wireless network, a user may run a test over Wi-Fi, then reposition their Wi-Fi AP and run the test again. These measurements may help the user optimize the placement of the wireless access point but, by design, they reflect the performance of the user's home wireless network, not that of the ISP. Tests that are user-initiated ("crowdsourced") are more likely to suffer from self-selection bias. It can be difficult to use these results to draw conclusions about an ISP, geographic region, and so forth.

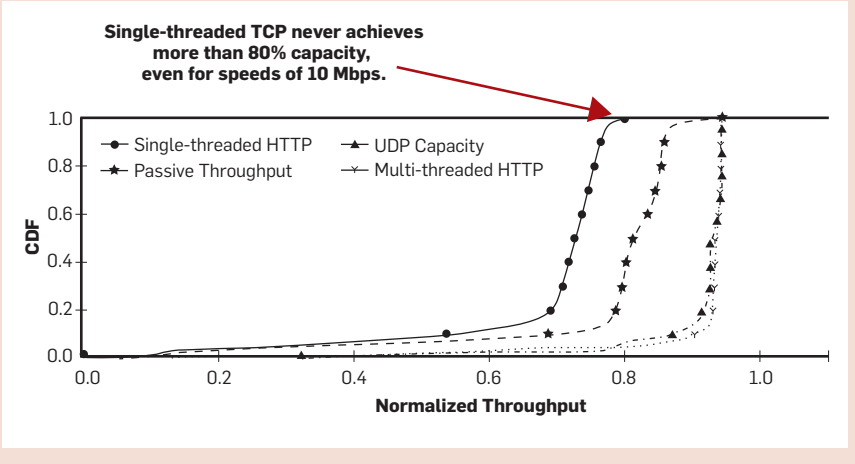
Infrequent testing. If tests are too infrequent or are only taken at certain times of day, the resulting measurements may not accurately reflect a user's Internet capacity. An analogy would be looking out a window once per day in the evening, seeing it was dark outside, and concluding that it must be dark 24 hours a day. Additionally, if the user only conducts a test when there is a transient problem, the resulting measurement may not be representative of the performance that a user typically experiences. Automatic tests run multiple times per day at randomly selected times during peak and off-peak times can account for some of these factors.

The Future of Speed Testing

Speed testing tools will need to evolve as end user connections approach and exceed 1Gbps, especially given that so many policy, regulatory, and investment decisions are based on speed measurements. As access network speeds increase and the performance bottlenecks move elsewhere on the path, speed test design must evolve to keep pace with both faster network technology and evolving user expectations. We recommend the following:

Retire outmoded tools such as NDT. NDT, also known as the Internet Health Test,[15] may appear at first glance to be suitable for speed tests. This is not the case, though it continues to be used for speed measurement despite its unsuitability and proven inaccuracy.[10] Its inadequacy for measuring access link

Figure 7. Throughput vs. number of TCP threads.



speeds has been well-documented.[2] One significant problem is that NDT still uses a single TCP connection, nearly two decades after this was shown to be inadequate for measuring link capacity. NDT is also incapable of reliably measuring access link throughput for speeds of 100Mbps or more, as we enter an era of gigabit speeds. The test also includes the initial TCP slow start period in the result, leading to a lower value of average throughput than the link capacity can support in TCP steady state. It also faces all of the user-related considerations that we discussed previously. It is time to retire the use of NDT for speed testing and look ahead to better methods.

Use native, embedded, and dedicated measurement techniques and devices. Web-based tests (many of which rely on Javascript) cannot transfer data at rates that exceed several hundred megabits per second. As network speeds increase, speed tests must be "native" applications or run on embedded devices (for example, home router, Roku, Eero, and AppleTV) or otherwise dedicated devices (for example, Odroid, Raspberry Pi, SamKnows "white box," and RIPE Atlas probes).

Control for factors along the end-to-end path when analyzing results. As we outlined earlier, many factors can affect the results of a speed test other than the capacity of the ISP link—ranging from cross-traffic in the home to server location and provisioning. As access ISP speeds increase, these limiting factors become increasingly important, as bottlenecks elsewhere along the end-to-end path become increasingly prevalent.

Measure to multiple destinations. As access network speeds begin to approach and exceed 1Gbps, it can be difficult to identify a single destination and end-to-end path that can support the capacity of the access link. Looking ahead, it may make sense to perform active speed test measurements to multiple destinations simultaneously, to mitigate the possibility that any single destination or end-to-end network path becomes the network bottleneck.

Augment active testing with application quality metrics. In many cases, a user's experience is not limited by the access network speed, but rather the performance of a particular application (for example, streaming video) under the available network conditions. As previously mentioned, even the most demanding streaming video applications require only tens of megabits per second, yet user experience can still suffer as a result of application performance glitches, such as changes in resolution or rebuffering. As access network speeds increase, it will be important to monitor not just "speed testing" but also to develop new methods that can monitor and infer quality metrics for a variety of applications.

Adopt standard, open methods to facilitate better comparisons. It is currently very difficult to directly compare the results of different speed tests, because the underlying methods and platforms are so different. Tools that select the highest result of several sequential tests, or the average of several, or the average of several tests after the highest and lowest have been discarded. As the FCC has stated:[13] "A well-documented, public methodology for tests is critical to understanding measurement results." Furthermore, tests and networks should disclose any circumstances that result in the prioritization of speed test traffic.

Beyond being well-documented and public, the community should also come to agreement on a set of standards for measuring access link performance and adopt those standards across test implementations. C

References

1. Apple: Resolve Wi-Fi and Bluetooth Issues Caused by Wireless Interference, 2019. https://support.apple.com/en-us/ HT201542.
2. Bauer, S., Clark, D.D., and Lehr, W. Understanding Broadband Speed Measurements. In *Technology Policy Research Conference* (*TPRC*), 2010.
3. Bufferbloat. https://www.bufferbloat.net.
4. CALSPEED Program, 2019; http://cpuc.ca.gov/General. aspx?id=1778.
5. *Electronic Engineering Times.* Avoiding iInterference in the 2.4-GHz ISM band, 2006; https: //www.eetimes.com/document.asp?doc_id=1273359.
6. FCC. Measuring Broadband America Program (Fixed), GN Docket No. 12264, Aug. 2013; https://ecfsapi.fcc.gov/file/ 7520939594.pdf.
7. FCC: Measuring Broadband America Program; https://www.fcc. gov/general/measuring-broadband-america.
8. FCC. Measuring Broadband America. Technical report, 2011; https://transition.fcc.gov/ cgb/measuringbroadbandreport/Measuring_U.S._-_Main_Report_Full.pdf.
9. FCC. MBA Report 2014, 2014; https://www.fcc.gov/ reports-research/reports/measuring-broadbandamerica/measuring-broadband-america-2014.
10. FCC. Letter to FCC on Docket No. 17-108, 2014; https://ecfsapi.fcc.gov/file/1083088362452/fcc-17-108-reply-aug2017.pdf.
11. FTP Rate Limiting, 2019; https://forum.filezillaproject.org/viewtopic.php?t=25895.
12. Gettys, J. Bufferbloat: Dark Buffers in the Internet. In IEEE Internet Computing, 2011.
13. Google Group. MLab speed test is incorrect? 2018; https://groups.google.com/a/measurementlab.net/forum/#!topic/discuss/vOTs3rcbp38.
14. Heckman, J. J. Selection bias and self-selection. In Econometrics, pages 201–224. 1990
15. Internet Health Test, 2019. http://internethealthtest.org/.
16. Lifewire. Change the Wi-Fi channel number to avoid interference, 2018; https://www.lifewire.com/Wi-Fi-channel-numberchange-to-avoid-interference-818208.
17. Livingood, J. Measurement Challenges in the Gigabit Era, June 2018. https://blog.apnic.net/2018/06/21/measurementchallenges-in-the-gigabit-era/.
18. Minnesota.gov. CheckspeedMN; https://mn.gov/deed/programsservices/broadband/checkspeedmn.
19. New York Attorney General's Office. A.G. Schneiderman Encourages New Yorkers To Test Internet Speeds And Submit Results As Part Of Ongoing Investigation Of Broadband Providers, 2017; https://ag.ny.gov/press-release/agschneiderman-encourages-new-yorkers-testinternet-speeds-and-submit-results-part.
20. New York Attorney General's Office. Are You Getting the Internet Speeds You Are Paying For? https: // ag.ny.gov/SpeedTest.
21. New York State Broadband Program Office-Speed Test; https: //nysbroadband.ny.gov/speed-test.
22. Nginx Rate Limiting, 2019; https://www.nginx.com/blog/ rate-limiting-nginx/.
23. Ofcom. Broadband speeds: Research on fixed line home broadband speeds, mobile broadband performance, and related research; https: //www.ofcom.org.uk/research-and-data/telecomsresearch/broadband-research/broadband-speeds.
24. Ookla Speedtest, 2019; https://speedtest.net/.
25. Oracle IaaS Pricing, 2019; https://cloud.oracle.com/en_US/iaas/pricing.
26. *PC World.* Six things that block your Wi-Fi, and how to fix them, 2011; https://www.pcworld.com/article/227973/six_ things_that_block_your_Wi-Fi_and_how_to_fix_ them.html.
27. Penn State University. A Broadband Challenge: Reliable broadband internet access remains elusive across Pennsylvania, and a Penn State faculty member is studying the issue and its impact, 2018; https://news.psu.edu/story/525994/2018/06/28/research/broadband-challenge.
28. Revolution Wi-Fi. The 2.4 GHz Spectrum congestion problem and AP form factors, 2015; http://www.revolutionWi-Fi.net/ revolutionWi-Fi/2015/4/the-dual-radio-apform-factor-is-to-blame-for-24-ghz-spectrumcongestion.
29. Sundaresan, S., De Donato, W., Feamster, N., Teixeira, R., Crawford, S., and Pescape, A. broadband Internet performance: A view from the gateway. In *Proceedings of ACM SIGCOMM* (Aug. 2011), 134–145.
30. Sundaresan, S., Feamster, N., and Teixeira, R. Home network or access link? Locating last-mile downstream throughput bottlenecks. In *Proceedings of the Intern. Conf. on Passive and Active Network Measurement* (2016), 111–123.
31. Tanenbaum, A.S. et al. *Computer Networks.* Prentice Hall, 1996.
32. Wehe, 2019; https://dd.meddle.mobi.
33. Xfinity Internet Minimum System Recommendations; https://www.xfinity.com/support/articles/requirementsto-run-xfinity-internet-service.

Nick Feamster (feamster@uchicago.edu) is the Neubauer Professor in the Computer Science Department and Director of the Center for Data and Computing at the University of Chicago, IL, USA.

Jason Livingood (jason_livingood@comcast.com) is Vice President at Comcast, Philadelphia, PA, USA.

Watch the authors discuss this work in the exclusive *Communications* video. https://cacm.acm.org/videos/measuring-internet-speed