



Measuring Home Broadband Performance

Srikanth Sundaresan, Walter de Donato, Nick Feamster, Renata Teixeira, Sam Crawford, Antonio Pescapé

► To cite this version:

Srikanth Sundaresan, Walter de Donato, Nick Feamster, Renata Teixeira, Sam Crawford, et al.. Measuring Home Broadband Performance. Communications of the ACM, 2012, 55 (11), pp.100-109. hal-00835036

HAL Id: hal-00835036

<https://hal.sorbonne-universite.fr/hal-00835036>

Submitted on 17 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Measuring Home Broadband Performance

Srikanth Sundaresan
Georgia Tech
Atlanta, USA
srikanth@gatech.edu

Renata Teixeira
CNRS/UPMC Sorbonne Univ.
Paris, France
renata.teixeira@lip6.fr

Walter de Donato
University of Napoli Federico II
Napoli, Italy
walter.dedonato@unina.it

Sam Crawford
SamKnows
London, UK
sam@samknows.com

Nick Feamster
Georgia Tech
Atlanta, USA
feamster@cc.gatech.edu

Antonio Pescapè
University of Napoli Federico II
Napoli, Italy
pescapè@unina.it

ABSTRACT

We present the results from the first study of Internet access link performance measured directly from home routers. In conjunction with the Federal Communication Commission's study of broadband Internet access in the United States, we investigate the throughput and latency of network access links from about 4,000 routers across eight ISPs. Our findings provide a snapshot of access network performance across the United States, offer insights on how access network performance should be measured and presented to users, and inform various ongoing efforts to evaluate the performance of access networks around the world.

1. INTRODUCTION

Of nearly two billion Internet users worldwide, about 500 million are residential broadband subscribers [12]. Broadband penetration will continue to increase, with people relying on home connectivity for day-to-day and even critical activities. Accordingly, the Federal Communication Commission (FCC) is developing performance-testing metrics for access providers [4, 10, 23]. Policymakers, home users, and Internet Service Providers (ISPs) need better ways to benchmark broadband Internet performance.

Unfortunately, benchmarking home Internet performance is not as simple as running one-time "speed tests". There exist countless tools to measure Internet performance [5, 6, 18, 19, 21]. Previous work has studied the typical download and upload rates of home access networks [8, 16]; others have found that modems often have large buffers [16], and that DSL links often have high latency [17]. These studies have shed some light on access-link performance, but they have typically run measurements either from an end-host inside the home (from the "inside out") or from a server on the wide-area Internet (from the "outside in"). Because these tools run from end-hosts, they cannot analyze the effects of confounding factors such as home network cross-traffic, the wireless network, or end-host configuration. Also, many of these tools run as one-time measurements and, without continual measurements of the same access link, cannot establish a baseline performance level or observe how performance varies over time.

We measure broadband Internet performance directly from the router that is connected to the user's ISP. Measuring the access link from the home router offers several advantages over conventional methods. First, the home router is typically always on. Second, because it connects the home network to the ISP's network (as shown

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2012 ACM 0001-0782/08/0X00 ...\$5.00.

Factor	How we address it
Wireless Effects	Use a wired connection to the modem.
Cross Traffic	Measure cross traffic and avoid it/account for it.
Router load	Use a well-provisioned router.
Server location	Choose a nearby server.
End-to-end path	Focus on characterizing the last mile.
Router configuration	Test configuration in practice and controlled settings.

Table 1: Confounding factors and how we address them.

ISP	Technology	SamKnows		BISmark Total
		Total	Active	
Comcast	Cable	864	560	4
AT&T	DSL/FTTN	787	335	10
TimeWarner	Cable	690	381	-
Verizon	DSL/FTTP	551	256	-
Cox	Cable	381	161	-
Qwest	DSL/FTTN	265	117	-
Charter	Cable	187	51	-
Cablevision	Cable	104	53	-

Table 2: The SamKnows and BISmark deployments. *Active* deployments are those that report more than 100 download throughput measurements over the course of our study.

in Figure 1) taking measurements from this vantage point allows us to control the effects of many confounding factors, such as the home wireless network and load on the measurement host. Table 1 summarizes the challenges involved in measuring access ISP performance, and how performing measurements from the home router addresses each of them.

We used two complementary deployments. First, we collected data from the routers in over 4,200 homes across the United States and eight different ISPs from a deployment sponsored by the Federal Communications Commission and operated by SamKnows. Second, we collected data from 16 homes in the BISmark deployment, spanning three ISPs in Atlanta. The SamKnows deployment provides a large user base, as well as diversity in ISPs, service plans, and geographical locations. We can access BISmark routers remotely and run repeated experiments to investigate the effect of factors that we could not study in a larger deployment. For example, to study the effect of modem choice on performance, we installed different modems in the same home and conducted experiments in the same controlled setting. Both deployments run a comprehensive suite of measurement tools that periodically measure throughput, latency, packet loss, and jitter. We use active measurement data from both deployments from December 14, 2010 to January 14, 2011. Table 2 lists the ISPs that we study and the number of measured access links for each of them.

We characterize access network throughput (Section 3) and latency (Section 4) from the SamKnows and BISmark deployments. We explain how our throughput measurements differ from common

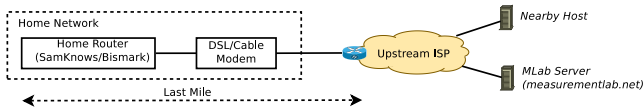


Figure 1: The home router sits directly behind the modem in the home network. It takes measurements both to the *last mile router* (first non-NAT IP hop on the path) and to wide area hosts.

“speed tests” and also propose several metrics that capture different aspects of latency. When our measurements cannot fully explain the observed behavior, we model the access link and test our hypotheses using controlled experiments. We find that the most significant factors affecting throughput are the access technology, ISPs’ traffic shaping policies, and congestion during peak hours. On the other hand, latency is mostly affected by the quality of the access link, modem buffering, and cross-traffic within the home.

This study offers many insights into both access network performance and the appropriate measurement methods for benchmarking home broadband performance, including:

- Throughput tends to be more variable during peak hours.
- ISPs use different policies and traffic shaping behavior that can make it difficult to compare measurements across them.
- Home network equipment and infrastructure can affect performance. For example, the amount of buffering in a user’s modem varies across models and can affect the latency that a user experiences during an upload (also known as the *bufferbloat* problem).
- There is no “best” ISP for all users. Some ISPs have better short-term throughput, while others may have better sustained throughput, lower latency, or generally more consistent performance. Different users may prefer different ISPs depending on their usage profiles and how those ISPs perform along performance dimensions that matter to them.

As the first in-depth analysis of home access network performance, our study offers insights for users, ISPs, and policymakers.¹ Users and ISPs can better understand the performance of the access link, as measured directly from the router; ultimately, such a deployment could help an ISP differentiate performance problems within the home from those on the access link. Our study also informs policy by illustrating that a diverse set of network metrics ultimately affect the performance that a user experiences. The need for a benchmark is clear, and the results from this study can serve as a principled foundation for such an effort.

2. ACCESS TECHNOLOGIES

We describe the two most common access technologies from our deployments: Digital Subscriber Line (DSL) and cable. A few users in our deployments have Fiber-To-The-Node (FTTN), Fiber-To-The-Premises (FTTP), and WiMax, but we do not have enough users to analyze these technologies.

DSL networks use telephone lines; subscribers have dedicated lines between their own DSL modems and the closest DSL Access Multiplexer (DSLAM). The DSLAM multiplexes data between the access modems and upstream networks, as shown in Figure 2a. The most common type of DSL access is asymmetric (ADSL), which

¹The original version of this paper “Broadband Internet Performance: A View from the Gateway” was published in the Proceedings of the ACM SIGCOMM 2011 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications.

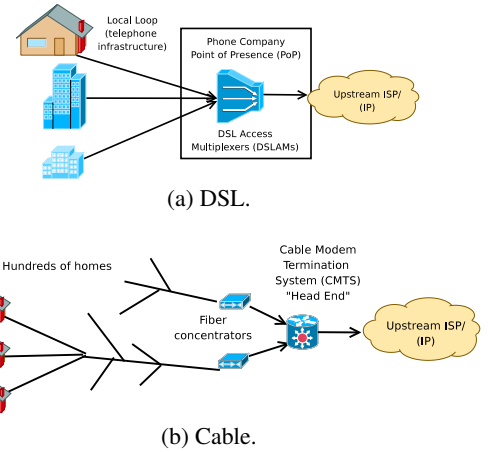


Figure 2: Access network architectures.

provides different upload and download rates. In cable access networks, groups of users send data over a shared medium (typically coaxial cable); at a regional *headend*, a Cable Modem Termination System (CMTS) receives these signals and converts them to Ethernet, as shown in Figure 2b. The physical connection between a customer’s home and the DSLAM or the CMTS is often referred to as the *local loop* or *last mile*. Users buy a service plan from a provider that typically offers some *maximum* capacity in both the upload and download directions.

ADSL capacity. The ITU-T standardization body establishes that the achievable rate for ADSL 1 [13] is 12 Mbits/s downstream and 1.8 Mbits/s upstream. The ADSL2+ specification [14] extends the capacity of ADSL links to at most 24 Mbits/s downstream and 3.5 Mbits/s upstream. Although the ADSL technology can theoretically achieve these speeds, many factors limit the capacity in practice. An ADSL modem negotiates the operational rate with the DSLAM (often called the *sync rate*); this rate depends on the quality of the local loop, which is mainly determined by the distance to the DSLAM from the user’s home and noise on the line. The maximum IP link capacity is lower than the sync rate because of the overhead of underlying protocols. The best service plan that an ADSL provider advertises usually represents the rate that customers can achieve if they have a good connection to the DSLAM. Providers also offer service plans with lower rates and can rate-limit a customer’s traffic at the DSLAM.

Modem configuration can also affect performance. ADSL users or providers configure their modems to operate in either *fastpath* or *interleaved* mode. In *fastpath* mode, data is exchanged between the DSL modem and the DSLAM in the same order that they are received, which minimizes latency but prevents error correction from being applied across frames. Thus, ISPs typically configure fastpath only if the line has a low bit error rate. *Interleaving* increases robustness to line noise at the cost of increased latency by splitting data from each frame into multiple segments and interleaving those segments with one another before transmitting them.

Cable capacity. In cable networks, the most widely deployed version of the standard is Data Over Cable Service Interface Specification version 2 (DOCSIS 2.0) [15], which specifies download rates up to 42.88 Mbits/s and upload rates up to 30.72 Mbits/s in the United States. The latest standard, DOCSIS 3.0, allows for hundreds of megabits per second by bundling multiple channels. Cable

Parameter	Type	Prot.	Freq.	Comments
SamKnows				
Downstream Throughput	Multi-threaded HTTP	TCP	2 hours	MLab, idle link
Upstream Throughput	Multi-threaded HTTP	TCP	2 hours	MLab, idle link
BISmark				
Downstream Throughput	Single-thread HTTP Passive throughput Capacity	TCP N/A UDP	30 min 30 min 12 hrs	curlget from Host /proc/net/dev ShaperProbe
Upstream Throughput	Single-thread HTTP Passive throughput Capacity	TCP N/A UDP	30 min 30 min 12 hrs	curlput to Host /proc/net/dev ShaperProbe

Table 3: SamKnows and BISmark throughput measurements.

providers often offer service plans with lower rates. An operator configures the service plan rate limit at the cable modem using a token bucket rate shaper. Many cable providers offer *PowerBoost*, which allows users to download (and, in some cases, upload) at rates that are higher than the contracted ones, for an initial part of a transfer. The actual rate that a cable user receives will vary with the network utilization of other users connecting to the same headend.

3. UNDERSTANDING THROUGHPUT

We first explore how different techniques for measuring throughput can generate different results and offer guidelines on how to interpret them. We then investigate the throughput users achieve on different access links. Finally, we explore the effects of ISP traffic shaping and the implications it has for throughput measurement.

3.1 How to Measure and Interpret Throughput

Users are often interested in the throughput that they receive on uploads or downloads, yet the notion of “throughput” can vary depending on how, when, and who is measuring it. For example, a run of `www.speedtest.net` in an author’s home, where the service plan was 6 Mbits/s down and 512 Kbits/s up, reported a downlink speed of 4.4 Mbits/s and an uplink speed of 140 Kbits/s. Netalyzr [16] reported 4.8 Mbits/s and 430 Kbits/s. Long-term measurements from the SamKnows router paint a different picture: the user achieves 5.6 Mbits/s down and 452 Kbits/s up. Both `www.speedtest.net` and Netalyzr measurements reflect transient network conditions, as well as other confounding factors.

There is no standard way to measure throughput. Bauer *et al.* list several notions of “broadband speed”: *capacity* is the total carrying capacity of the link; and the *bulk transfer capacity* is the amount of data that can be transferred along a path with a congestion-aware protocol like TCP [3]. The SamKnows routers measure bulk transfer capacity using three parallel HTTP transfers; this approach increases the likelihood of saturating the access link. The software first executes a “warmup” transfer until throughput is steady to ensure that the throughput measurements are not affected by TCP slow start. The download tests that follows use the same TCP connection to exploit the “warmed up” session. The tests last for about 30 seconds; the software reports snapshots of how many bytes were transferred for every five-second interval. BISmark measures throughput by performing an HTTP download and upload for 15 seconds using a single-threaded TCP connection once every 30 minutes, regardless of cross-traffic. To account for cross-traffic, the router counts bytes transferred by reading from `/proc/net/dev`, and compute the “passive throughput” as the byte count after the HTTP transfer minus the byte count before the transfer, divided by the transfer time. Table 3 summarizes the throughput measurements collected by the two deployments. Although measuring throughput may seem straightforward, our results demonstrate the extent to which different measurement meth-

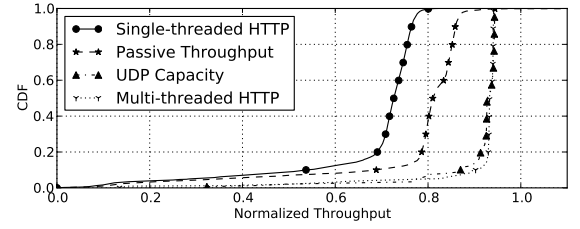


Figure 3: Comparison of various methods of measuring throughput. (SamKnows and BISmark)

ods can produce different results and, hence, may result in different conclusions about an ISP’s performance.

Throughput measurement techniques—even commonly accepted ones—can yield variable results. We perform comparisons of throughput measurement techniques in two locations that have both the SamKnows and BISmark routers (we use only two locations due to the logistical difficulty in deploying both routers in the same location). In both cases, the ISP is AT&T, but the service plans are different (6 Mbits/s down and 512 Kbits/s up; and 3 Mbits/s down and 384 Kbits/s up). We normalize the throughput with the advertised service plan so we can compare the service plans. Figure 3 shows a CDF of the normalized throughput reported by the four methods from Table 3. Each data point represents a single throughput measurement. A value of 1.0 on the x-axis indicates that the throughput matches the ISP’s advertised rate; no method achieves that value. This could be due to many factors: the sync rate of the modem to the DSLAM; layer-2 framing overhead on the line; or overhead from the measurement techniques themselves. Multiple parallel TCP sessions nearly achieve the advertised throughput. UDP measurements also produce consistent measurements of throughput that are closer to the multi-threaded TCP measurement. A single-threaded TCP session may not achieve the same throughput, but accounting for cross-traffic with passive measurements does yield a better throughput estimate.

The behavior of single-threaded TCP measurements varies for different access links. We compare throughput for two BISmark users with the *same ISP and service plan* (AT&T; 3 Mbits/s down, 384 Kbits/s up) who live only a few blocks apart. Figure 4 shows that User 2 consistently sees nearly 20% more throughput than User 1. One possible explanation for this difference is the loss rates experienced by these two users: User 1 experiences four times more packet loss in both directions than User 2. The baseline latencies also differ by about 16 milliseconds (8 ms vs. 24 ms). We confirmed from the respective modem portals that User 1 has interleaving disabled and User 2 has interleaving enabled. Thus, User 2’s connection recovers better from line noise. Single-threaded downloads suffer more from high packet loss rates than multi-threaded downloads; interleaving reduces the packet loss rate, thus improves the performance of a single-threaded download. For the rest of the paper, we consider only multi-threaded TCP throughput.

Takeaway: Different throughput measurement techniques capture different aspects of throughput. A single-threaded TCP session is sensitive to packet loss. Augmenting this measurement with passive usage measurements improves its accuracy. Multi-threaded TCP and the UDP capacity measurements measure the access link capacity more accurately and are more robust to loss.

3.2 Throughput Performance

We investigate the throughput obtained by users in the Sam-

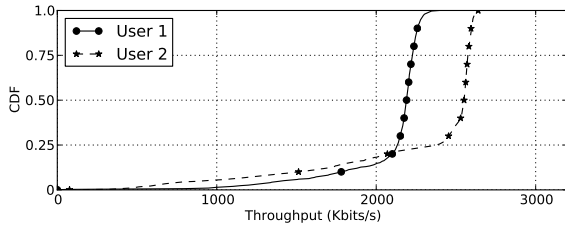


Figure 4: Users with the same service plan but different loss profiles see different performance. User 1 has higher loss and sees lower performance. (BISmark)

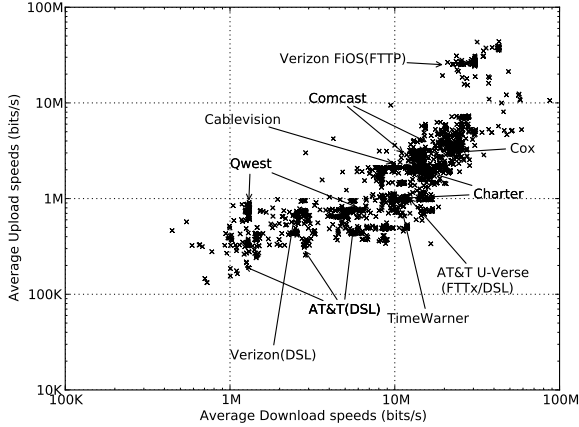


Figure 5: Average download rate versus the average upload rate obtained by individual users in the dataset. (SamKnows)

Knows deployment and the consistency of their performance.

What performance do users achieve? Figure 5 shows the average download and upload speeds obtained by each user in the SamKnows dataset. Each point in the scatterplot shows the average performance obtained by a single user in the deployment. Clusters of points in the plot reveal common service plans of different ISPs, identified in the plot by labels. In general, these results agree with previous [8, 16] work, although our dataset also includes Verizon FiOS (FTTP) users that clearly stand out and other recent service offerings (*e.g.*, AT&T U-Verse). Although there are some noticeable clusters around various service plans, there appears to be considerable variation even within a single service plan. We seek to characterize both the performance variations and their causes.

Do users achieve consistent performance? We analyze how consistently users in the SamKnows deployment achieve their peak performance using the $Avg/P95$ metric, which we define as the ratio of the average upload or download throughput obtained by a user to the 95th percentile of the upload or download throughput value obtained by the same user. Higher values for these ratios reflect that users’ upload and download rates are more consistently close to the highest rates that they achieve; lower values indicate that user performance fluctuates.

Figure 6 shows the CDF of the $Avg/P95$ metric across users of each ISP. Most access links achieve throughput close to their 95th percentile value. Certain ISPs (*e.g.*, Cox, Cablevision) achieve an average download throughput that is significantly less than their 95th percentile. Upload throughput is much more consistent, possibly because upload rates are typically much lower.

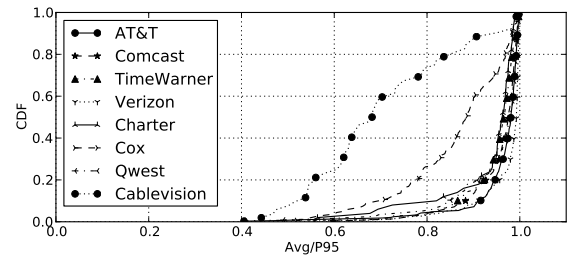


Figure 6: Consistency of throughput performance: The average throughput of each user is normalized by the 95th percentile value obtained by that user. (SamKnows)

Why is performance sometimes inconsistent? One possible explanation for inconsistent download performance is that the access link may exhibit different performance characteristics depending on time of day. Figure 7a shows the $Avg/P95$ metric across the time of day. We obtain the average measurement reported by each user at that time of day and normalize it with the 95th percentile value of that user over all reports. Cablevision users see, on average, a 40% drop in performance in the peak evening hours; Cox users experience a 20% performance reduction on average. This effect exists for other ISPs to a lesser extent, confirming prior findings [8]. Without knowing the service plan for each user, we cannot, in general, say whether the decrease in performance represents a drop below the service plans for those users (*e.g.*, these users might see rates higher than their plan during off-peak hours). However, the FCC’s report [1], which analyzes performance in comparison to advertised rates, confirms that Cablevision users do see average performance significantly less than the advertised rates. Figure 7b shows how the standard deviation of normalized throughput varies depending on the time of day. All ISPs experience more variable performance during peak hours. Most ISPs do not see an increase in loss rates during peak hours, but Cox does. ISPs that exhibit poor performance during peak hours may be under-provisioned; they may be experiencing congestion or explicitly throttling traffic during peak hours.

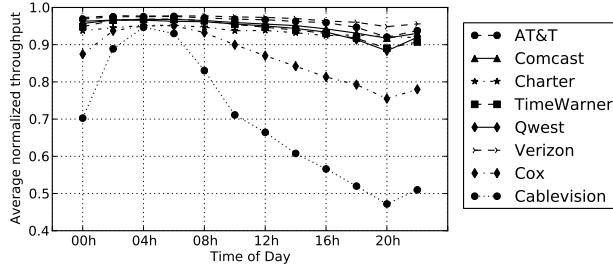
Takeaway: Throughput performance is more variable during peak hours. A one-time “speed test” measurement taken at the wrong time could likely report misleading numbers that do not have much bearing on performance over time.

3.3 Effect of Traffic Shaping on Throughput

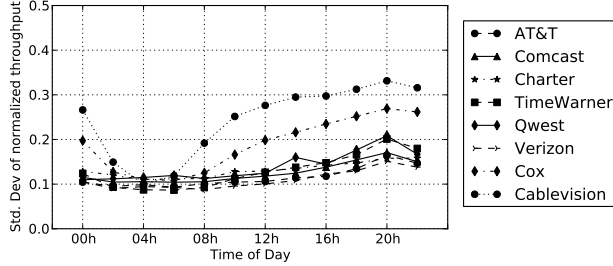
ISPs shape traffic in different ways, which makes it difficult to compare measurements across ISPs, and sometimes even across users within the same ISP. We study the effect of PowerBoost across different ISPs, time, and users. We also model how Comcast implements PowerBoost.

Which ISPs use PowerBoost, and how does it vary across ISPs?

Each SamKnows throughput measurement lasts 30 seconds, and each report is divided into six snapshots at roughly five-second intervals for the duration of the test. This technique highlights the evolution of throughput performance over time. On a link that is subject to traffic shaping, the throughput during the last snapshot will be less than the throughput during the first. For each report, we normalize the throughput in each period by the throughput reported for the first period. The normalized throughput on an unshaped link is close to one for all intervals. On the other hand, on an access link configured with PowerBoost, the throughput in the last five seconds should be less than the throughput in the first



(a) Peak and worst performance differ by upto 40%.



(b) The standard deviation of throughput measurements increases during peak hours, most significantly for ISPs that see lower throughputs at peak hours.

Figure 7: Time of day is significant: The average download throughput for Cablevision and Cox users drops significantly during the evening peak time. Throughput is also significantly more variable during peak time. (SamKnows)

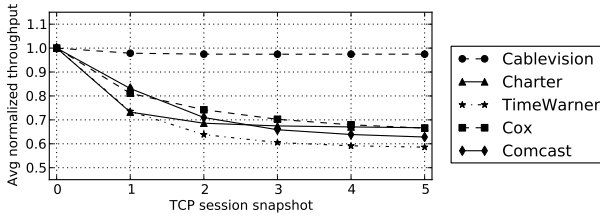


Figure 8: The average throughput during the the measurement decreases for the ISPs that enable PowerBoost. (SamKnows)

five seconds (assuming that PowerBoost lasts less than 30 seconds). Figure 8 shows the average progression of throughput over all users in an ISP: the average normalized throughput decreases over time. Our data shows that most cable ISPs provide some level of PowerBoost for less than 30 seconds, at a rate of about 50% more than the normal rate. Cablevision's line is flat; this suggests that either it does not provide PowerBoost, or it lasts well over 30 seconds consistently, in which case the throughput test would see only the PowerBoost effect. The gradual decrease, rather than an abrupt decrease, could be because PowerBoost durations vary across users or that the ISP changes PowerBoost parameters based on network state. In the case of uploads, only Comcast and Cox seem to deploy PowerBoost; in these cases we observed a difference in throughput of about 20%. DSL ISPs do not appear to implement PowerBoost.

Do different users see different PowerBoost effects? We investigate Comcast's use of PowerBoost using the BISmark testbed. According to Comcast [7], their implementation of PowerBoost provides higher throughput for the first 10 MBytes of a download and the first 5 MBytes of an upload. We measure the shaped throughput for download and upload at the receiver using `tcpdump`. Because

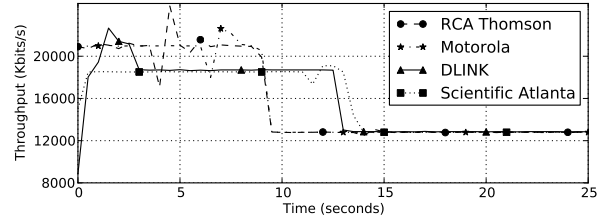


Figure 9: The level and duration of the burstiness is different for users with different modems, suggesting different shaping mechanisms or parameters. (BISmark)

Parameter	Type	Prot.	Freq.	Comments
SamKnows				
Latency	End-to-end	UDP	600 pkts/hr	MLab
	End-to-end	ICMP	5 pkts/hr	MLab
	Last-mile	ICMP	5 pkts/hr	First IP hop
	Upstream load	ICMP	2 hours	During upload
Loss	Downstream load	ICMP	2 hours	During download
	End-to-end	UDP	600 pkts/hr	MLab
Jitter	Bi-directional	UDP	1 hour	500pkts/30sec
Web GET	HTTP	TCP	1 hour	Alexa sites
BISmark				
Latency	End-to-end	ICMP	5 min	Host
	Last-mile	ICMP	5 min	First IP hop
	Upstream load	ICMP	30 min	During upload
	Downstream load	ICMP	30 min	During download
Packet loss	End-to-end	UDP	15 min	D-ITG
Jitter	End-to-end	UDP	15 min	D-ITG

Table 4: SamKnows and BISmark latency and loss tests.

our tests are intrusive, we conducted them only a few times, but the results are consistent across traffic generators and ports.

Figure 9 shows the downstream throughput profiles for four users, each identified by their modem type. Although the modem does not affect burst rates, they do have different amounts of buffering, which can affect latency. All four users experience PowerBoost effects, but, surprisingly, each user experiences a different traffic shaping profile: The user with a D-LINK modem sees a peak rate of about 21 Mbits/s for 3 seconds, 18.5 Mbits/s for a further ten seconds, and a steady-state rate of 12.5 Mbits/s. The Motorola user sees a peak rate of 21 Mbits/s for about 8 seconds. The D-LINK profile can be modeled as a cascaded token bucket filter with rates of 18.5 Mbits/s and 12.5 Mbits/s, and buffer sizes of 10MBytes and 1Mbyte respectively, with a capacity of 21 Mbits/s. Upload profiles vary across different users as well, although the shaping profiles seem to indicate that only a single token bucket is applied on the uplink.

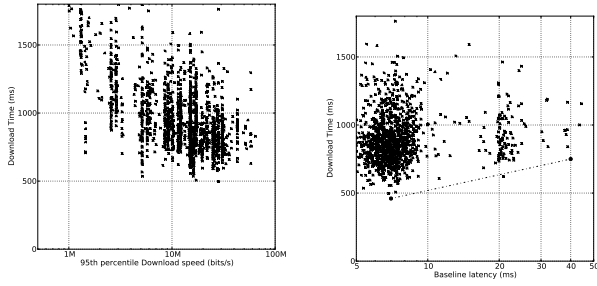
Takeaway: Many cable ISPs implement PowerBoost, which could distort speedtest-like measurements. In particular, any throughput measurement that lasts less than 35 seconds will mainly capture the effects of PowerBoost. While some people may be only interested in short-term burst rates, others may be more interested in long-term rates. Any throughput benchmark should aim to characterize both burst rates and steady-state throughput rates.

4. UNDERSTANDING LATENCY

We show how latency can drastically affect performance, even on ISP service plans with high throughput. We then study how various factors ranging from the user's modem to ISP traffic shaping policies can affect latency.

4.1 How (and Why) to Measure Latency

Latency not only affects the throughput that users achieve, it



(a) Fetch time stabilizes above 6Mbits/s. (b) Latency affects fetch times.

Figure 10: Effect of downstream throughput and baseline latency on fetch time from `facebook.com`. (SamKnows)

also affects the performance that users perceive, for many applications, since it affects everything from DNS lookup time to the time to set up a TCP connection. Although measuring latency appears straightforward, arriving at the appropriate metric is a subtle challenge because our goal is to isolate the performance of the access link from the performance of the end-to-end path. End-to-end latency between endpoints is a common metric in network measurement, but it reflects the delay that a user experiences along a wide-area path. We use two metrics that are more appropriate for access networks: The first metric is the *last-mile latency*, which is the latency to the first IP hop inside the ISP’s network. The last-mile latency captures the latency of the access link, which could affect gaming or short downloads. We measure last-mile latency in both of our deployments. The second metric we define is *latency under load*, which is the latency that a user experiences during an upload or download (*i.e.*, when the link is saturated in either direction). For BISmark, we measure the last-mile latency under load; on the SamKnows platform, we measure end-to-end latency under load. Table 4 summarizes the latency measurements we collect.

We investigate the effect of last-mile latency on download times for popular Web pages. Figure 10 shows the download time for `www.facebook.com` and how it varies by both the user’s throughput and baseline last-mile latency. Figure 10a plots the 95th percentile of each user’s downstream throughput versus the average time it takes to download all objects from `www.facebook.com`. The average size of the download is 125 KBytes. As expected, the download times decrease as throughput increases; interestingly, there is negligible improvement beyond a rate of 6 Mbits/s. Figure 10b plots download time against the baseline latency for all users whose downstream throughput (95th percentile) exceeds 6 Mbits/s. Minimum download times increase by about 50% when baseline latencies increase from 10 ms to 40 ms. The pronounced effect of latency on download time for Web objects underscores the influence of baseline latency.

4.2 Last-Mile Latency

We obtain the last-mile latency by running `traceroute` to a wide-area destination and extracting the first IP address along the path that is not a NAT address. Note that we are measuring the latency to the first network-layer hop, which may not in fact be the DSLAM or the CMTS, because some ISPs have layer-two DSLAMs that are not visible in `traceroute`. The possibility of measuring slightly further than the DSLAM or CMTS should not materially affect our results, since the latency between hops inside an

ISP is typically much smaller than the last-mile latency.

How does access technology affect last-mile latency? Last-mile latency is generally quite high, varying from about 10 ms to nearly 40 ms (ranging from 40–80% of the end-to-end path latency). Variance is also high. One might expect that variance would be lower for DSL, since it is not a shared medium like cable. Surprisingly, the opposite is true: AT&T and Verizon have high variance compared to the mean. Qwest also has high variance, though it is a smaller fraction of the mean. To understand this variance, we divide different users in each ISP according to their baseline latency, as shown in Figure 11. Most users of cable ISPs are in the 0–10 ms interval. On the other hand, a significant proportion of DSL users have baseline last-mile latencies more than 20 ms, with some users seeing last-mile latencies as high as 50 to 60 ms. Based on discussions with network operators, we believe DSL ISPs may be enabling an interleaved local loop for these users. We also analyzed loss and jitter characteristics. We saw that the average loss is small, but variance is high for all ISPs, suggesting bursty loss. Jitter has similar characteristics; while the average jitter is low, the variation is high, especially on the upstream. This is probably due to the asymmetric nature of access link throughputs.

Takeaway: Cable users tend to have lower last-mile latency and jitter, while for DSL users it may vary significantly based on physical factors such as distance to the DSLAM or line quality.

4.3 Latency Under Load

We turn our attention to “bufferbloat”, a problem that has gathered much interest recently because of its effects on many aspects of performance, including latency under load [11].

Problem: Bufferbloat. Buffers on DSL and cable modems are too large. Buffering affects latency during periods when the access link is loaded; during these periods, packets can see substantial delays as they queue in the buffer. The capacity of the uplink also affects the latency that buffering introduces. For a given buffer size, the queuing delay will be lower on access links with higher capacities because the draining rate is higher. We study the effect of buffering on access links by measuring latency when the access link is saturated, under the assumption that the last-mile is the bottleneck.

How widespread are oversized buffers? Figure 12 shows the average ratios of latency under load to baseline latency for each user across different ISPs for the SamKnows data. The histogram shows the latencies when the uplink and the downlink are saturated separately. This figure confirms that oversized buffers affect users across all ISPs, albeit to different extents. The factor of increase when the uplink is saturated is much higher than it is when the downlink is saturated. One plausible explanation is that the downlink usually has more capacity than the uplink, so buffering on the ISP side is lower. The home network is often better provisioned than the downlink, so downstream traffic experiences less buffering in the modem. The high variability in the latency under load can be partly explained by the variety in service plans; for instance, AT&T offers plans ranging from 768 Kbits/s to 6 Mbits/s for DSL and up to 18 Mbits/s for UVerse and from 128 Kbits/s to more than 1 Mbit/s for upstream. In contrast, Comcast offers fewer service plans, which makes it easier to design a device that works well for all service plans.

How does modem buffering affect latency under load? To study the effects of modem buffers on latency under load, we conduct tests on AT&T and Comcast modems using BISmark. We ran tests on the best AT&T DSL plan (6 Mbits/s down; 512 Kbits/s up). We

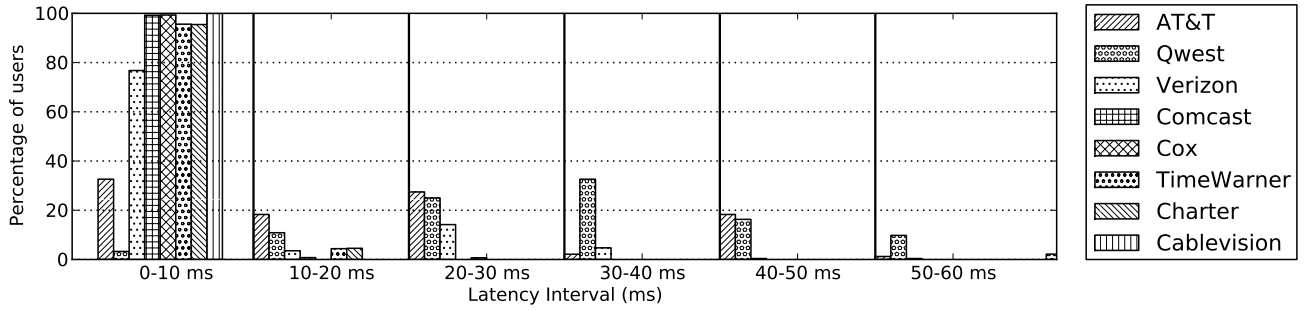


Figure 11: The baseline last mile latency for each user is computed as the 10^{th} percentile of the last mile latency. Most users see latencies less than 10 ms, but there are a significant number of users with the last mile latency greater than 10 ms. (SamKnows)

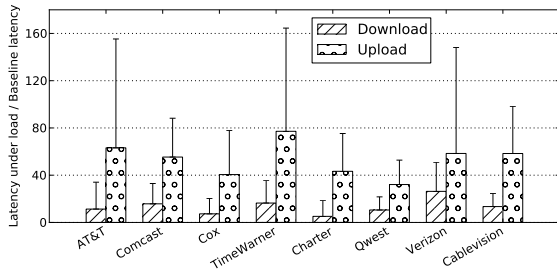


Figure 12: Latency under load: the factor by which latency goes up when the upstream or the downstream is busy translate to significant real latencies, often in the order of seconds. (SamKnows)

first started ICMP ping to the last mile hop. After 30 seconds, we flooded the uplink (at 1 Mbits/s for AT&T and at 10 Mbits/s for Comcast, using `iperf`'s UDP measurement). After 60 seconds, we stopped `iperf`, but let ping continue for another 30 seconds. The ping measurements before and after the `iperf` test established the baseline latency.

Figure 13 shows the latency under load for three different DSL modems. In all cases, the latency skyrockets when flooding begins and plateaus when the buffer is saturated. This latency plateau indicates the size of the buffer, since we know the uplink draining rate. Surprisingly, we observed more than an order of magnitude of difference in buffering in different modems. The 2Wire modem introduces the lowest worst-case latency of 800 ms, the Motorola modem about 1600 ms, while the Westell introduces more than 10 seconds of latency! Comcast users experienced as much as 350 ms of latency under load. Because modems are usually the same across service plans, we expect that latency under load may be even worse for users with slower plans (and, hence, slower drain rates).

We perform experiments in Emulab [9] to model modem buffering; the topology has two end-hosts and one router. We configured a token bucket filter using `tc` with the buffer size computed as $512 \text{ Kbits/s} \times \max(\text{latency of modem})$. This yields 640 KBytes for Westell, 100 KBytes for Motorola, and 55 KBytes for 2Wire. This simple setup almost perfectly captures the latency profile that the actual modems exhibit. We observed little difference in throughput for the three buffer sizes. We also emulated other buffer sizes. For a 512 Kbits/s uplink, we observed that the modem buffers exceeding 20 KBytes do little for throughput, but cause a linear increase in latency under load. Our experiments confirm that buffer sizes in all three modems are too large for the uplink.

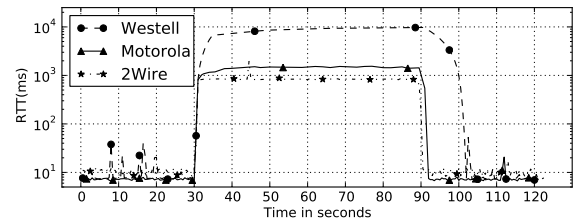


Figure 13: Different buffer sizes across modems lead to wide disparities in latencies when the upstream link is busy. (BISmark)

Can data transfer be modified to improve latency under load?

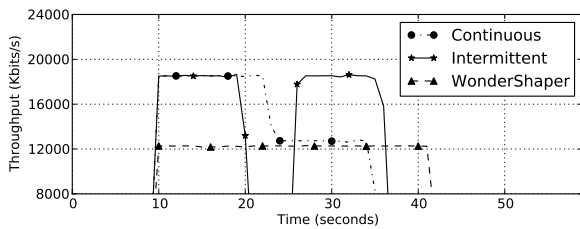
We explore how large “bulk” flows and delay-sensitive flows can co-exist without interfering with one another. We compare the effects of a 50 MByte download on a G.711 VoIP call in three different conditions: (1) not applying any traffic control, (2) sending intermittent traffic at capacity on 10.8 seconds ON and 5.3 seconds OFF cycle, and (3) shaping using the WonderShaper [24] tool. Figure 14 shows the result of this experiment. Without traffic control, the transfer takes 25.3 seconds; however, just after the PowerBoost period, the VoIP call starts suffering high latency and loss until the end of the transfer. In the second scenario, traffic is sent in pulses, and the download takes 26.9 seconds. In third case, traffic is sent at just under the long term rate and the download takes 32.2 seconds. Sending intermittent traffic and shaping the traffic with WonderShaper do not increase latency much, because they do not ever fully deplete the tokens. The appropriate ON/OFF periods for intermittent transfers depend on the token bucket parameters,² and the size of the file to be transferred. Both approaches achieve similar long-term rates but yield significant latency benefit. These approaches need the shaping parameters to be tuned to work properly, however.

Takeaway: Modem buffers are too large. The smallest buffers we see induce nearly one-second latency under load for AT&T and 300 ms for Comcast. Buffering degrades both interactivity and throughput. Transferring data in shorter bursts or shaping traffic using tools like WonderShaper can mitigate buffering problems.

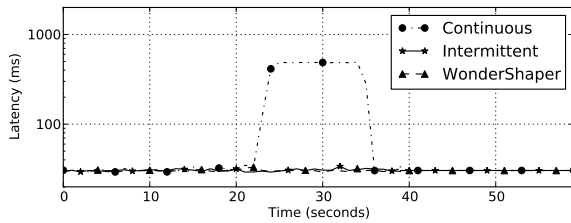
5. LESSONS LEARNED

We conclude with some high-level lessons and suggestions for

²If ρ_r is the rate we want to reserve for real-time applications, and ρ_t the token rate, the condition to be satisfied is: $(\rho_b + \rho_r - \rho_t) \times \tau_{on} \leq \tau_{off} \times (\rho_t - \rho_r)$, where ρ_b is the sending rate during the pulse, and τ_{on} and τ_{off} are the ON and the OFF times, respectively.



(a) Throughput.



(b) Latency.

Figure 14: It is possible to maintain low latency by modifying data transfer behavior. (BISmark)

future research directions. One significant takeaway for users, policymakers, ISPs, and researchers is that *understanding access network performance requires periodic measurement from the home router*. Existing speed tests and end-to-end latency measurements do not reflect access network performance over an extended period of time, and they neglect confounding factors within the home network. Our study of broadband networks yields several lessons:

Lesson 1 (One Measurement Does Not Fit All) *Different ISPs use different policies and traffic shaping behaviors that make it difficult to compare measurements across ISPs.*

There is no single number that characterizes performance, or even throughput. Certain ISP practices such as PowerBoost can distort benchmarking measurements; ISPs might even design their networks so that widely used performance tests yield good performance. Developing a benchmarking suite for ISP performance that users can understand (*e.g.*, in terms of the applications they use) is critical; the measurements we develop in this paper may be a good starting point for that. Along these lines, more work is needed to understand the performance of specific applications, such as how video streaming performance compares across ISPs. The Netflix study on ISP streaming performance [20] is a good start, but more such performance benchmarks are needed.

Lesson 2 (One ISP Does Not Fit All) *There is no “best” ISP for all users. Different users may prefer different ISPs depending on their usage profiles and how those ISPs perform along performance dimensions that matter to them.*

Different ISPs may be “better” along different performance dimensions, and the service plan that a user buys is only part of the picture. For example, we saw that, above a certain throughput, latency is the dominant factor in determining Web page load time. Similarly, a gamer might be interested in low latency or jitter, while an avid file swapper may be more interested in high throughput. An imminent technical and usability challenge is to summarize access network performance data so that users can make informed choices about the service plans that are most appropriate for them (akin to a

“performance nutrition label” [2]). Our recent work proposes some first steps in this direction [22].

Lesson 3 (Home Network Equipment Matters) *A user’s home network infrastructure can significantly affect performance.*

Modems can introduce latency variations that are orders of magnitude more than the variations introduced by the ISP. Other effects inside the home that we have not yet studied, such as the wireless network, may also ultimately affect the user’s experience. More research is needed to understand the characteristics of traffic inside the home and how it affects performance.

Acknowledgments

We thank the participants in the SamKnows and BISmark studies, and to Walter Johnston at the FCC for help and access to the data from the SamKnows study. This project is supported by the the National Science Foundation through awards CNS-1059350, CNS-0643974, a generous Google Focus Grant, the European Community’s Seventh Framework Programme (FP7/2007-2013) no. 258378 (FIGARO), and the ANR project C’MON.

6. REFERENCES

- [1] Measuring Broadband America: A Report on Consumer Wireline Broadband Performance in the U.S. http://www.fcc.gov/cgb/measuringbroadbandreport/Measuring_U.S._Main_Report_Full.pdf.
- [2] Does broadband need its own government nutrition label? <http://arstechnica.com/tech-policy/news/2009/10/does-broadband-needs-its-own-government-nutrition-label.ars>, Oct. 2010. Ars Technica.
- [3] S. Bauer, D. Clark, and W. Lehr. Understanding broadband speed measurements. In *38th Research Conference on Communication, Information and Internet Policy*, Arlington, VA, 2010.
- [4] K. Bode. FCC: One Million Speedtests and Counting. <http://www.dslreports.com/shownews/FCC-One-Million-Speedtests-And-Counting-109440>, July 2010.
- [5] A. Botta, A. Dainotti, and A. Pescapé. A tool for the generation of realistic network workload for emerging networking scenarios. *Computer Networks*, 2012.
- [6] R. Carlson. Network Diagnostic Tool. <http://e2epi.internet2.edu/ndt/>.
- [7] Comcast FAQ. <http://customer.comcast.com/Pages/FAQViewer.aspx?Guid=024f23d4-c316-4a58-89f6-f5f3f5dbdcdf6>, Oct. 2007.
- [8] M. Dischinger, A. Haeberlen, K. P. Gummadi, and S. Saroiu. Characterizing residential broadband networks. In *Proc. ACM SIGCOMM Internet Measurement Conference*, San Diego, CA, USA, Oct. 2007.
- [9] Emulab. <http://www.emulab.net/>, 2006.
- [10] National Broadband Plan. <http://www.broadband.gov/>.
- [11] J. Gettys. Bufferbloat. <http://www.bufferbloat.net/>.
- [12] Internet World Stats. <http://www.internetworldstats.com/dsl1.htm>.
- [13] Asymmetric Digital Subscriber Line Transceivers. ITU-T G.992.1, 1999.
- [14] Asymmetric Digital Subscriber Line (ADSL) Transceivers - Extended Bandwidth ADSL2 (ADSL2Plus). ITU-T G.992.5, 2003.
- [15] Data-over-cable service interface specifications: Radio-frequency interface specification. ITU-T J.112, 2004.
- [16] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzer: Illuminating the edge network. In *Proc. Internet Measurement Conference*, Melbourne, Australia, Nov. 2010.
- [17] G. Maier, A. Feldmann, V. Paxson, and M. Allman. On dominant characteristics of residential broadband internet traffic. In *ACM Internet Measurement Conference*, 2009.
- [18] M. Mathis. et al. Network Path and Application Diagnosis. <http://www.psc.edu/networking/projects/pathdiag/>.
- [19] Netalyzer. <http://netalyzer.icsi.berkeley.edu/>.
- [20] Netflix Performance on Top ISP Networks. <http://techblog.netflix.com/2011/01/netflix-performance-on-top-isp-networks.html>, Jan. 2011.
- [21] ShaperProbe. <http://www.cc.gatech.edu/~partha/diffprobe/shaperprobe.html>.
- [22] S. Sundaresan, N. Feamster, R. Teixeira, A. Tang, K. Edwards, R. Grinter, M. Chetty, and W. de Donato. Helping users shop for ISPs with internet nutrition labels. In *ACM SIGCOMM Workshop on Home Networks*, 2011.
- [23] D. Vorhaus. A New Way to Measure Broadband in America. <http://blog.broadband.gov/?entryId=359987>, Apr. 2010.
- [24] WonderShaper. <http://lartc.org/wondershaper/>, 2002.