The Impact of IP Version on Household Internet Speed: A Comparative Study

AMANDA HSU, Georgia Institute of Technology, USA
NICK FEAMSTER, University of Chicago, USA
PAUL PEARCE, Georgia Institute of Technology, USA
FRANK LI, Georgia Institute of Technology, USA

Despite extensive research measuring broadband quality, limited work has considered the interaction between residential Internet throughput and IP protocol. Public speed tests used to determine Internet quality do not, by and large, control for IP version; results, analysis, and recommendations are made using a hybrid of IPv4 and IPv6 data. Given the range of protocol designs, software stacks, and network infrastructure differences between the protocols, combined with the recent significant increase in IPv6 adoption, it is critical to understand what role the IP protocol plays in measuring Internet speeds.

In this work, we systematically compare IPv4 and IPv6 speeds in residential access networks, examining differences in throughput experienced by households depending on IP version. Our findings demonstrate that IPv4 and IPv6 throughput differ in many instances and motivate a large-scale re-evaluation of our assumptions on IP version in future speed test analysis. Specifically, we find that IPv4 and IPv6 speeds differ in a significant number of cases, with up to 18.3% of our measurements differing by over 5%. Our findings indicate that substantial speed differences between IP versions can be driven by provider-specific factors such as speed tiers. Furthermore, we observe differences in IPv4 and IPv6 data depending on the speed-test software and testing infrastructure. Thus, this work guides future research about Internet speeds on how to consider and control for IP version.

CCS Concepts: • **Networks** → **Network measurement**.

Additional Key Words and Phrases: Network measurement; Access networks; Throughput; IP protocol

ACM Reference Format:
Amanda Hsu, Nick Feamster, Paul Pearce, and Frank Li. 2025. The Impact of IP Version on Household Internet Speed: A Comparative Study. *Proc. ACM Meas. Anal. Comput. Syst.* 9, 3, Article 64 (December 2025), 28 pages. https://doi.org/10.1145/3771579

1 Introduction

Internet access is vital, arguably at the level of a basic human right [18]. Researchers and policy makers alike rely on *speed tests* to understand connectivity quality from the perspective of residential networks [10]. To this end, the United States Federal Communications Commission (FCC) publishes Internet quality measurements via the Measuring Broadband America (MBA) project, which for over 10 years has placed speed test probes in the homes of volunteers in order to evaluate key performance metrics for various Internet Service Providers (ISPs) [14]. This and other speed test datasets have been used to highlight a variety of inequalities by geographic location, service provider, and Customer Premise Equipment (CPE) [32, 42, 53, 56, 59].

Authors' Contact Information: Amanda Hsu, ahsu67@gatech.edu, Georgia Institute of Technology, USA; Nick Feamster, feamster@uchicago.edu, University of Chicago, USA; Paul Pearce, pearce@gatech.edu, Georgia Institute of Technology, USA; Frank Li, frankli@gatech.edu, Georgia Institute of Technology, USA.

ACM 2476-1249/2025/12-ART64
https://doi.org/10.1145/3771579

Despite extensive and impactful speed test measurements [11], one critical facet that has not been explored in depth is how speeds differ between IP versions. With IPv4 exhaustion, IPv6 deployment has increased dramatically. As of fall 2025, Google estimates that almost half of its traffic is IPv6 [17], with adoption having accelerated significantly since 2020. While IPv6 has allowed the Internet to continue expanding past IPv4 exhaustion [47], the Internet is now a hybrid of IPv4, IPv6, and dual-stack networks. Residential networks are no different and may use IPv4, IPv6, or both, depending on factors such as ISP, hardware, software, or service.

IPv4 and IPv6, while both IP protocols, are inherently incompatible and require translation to interoperate. Moreover, protocols have different software stacks and consumer hardware (e.g., hardware offload [4]). Most notably, IPv4 connections are likely to traverse a Network Address Translation (NAT) technology, implemented to extend the use of the IPv4 address space [48]. Such differences introduce distinct potential for bottlenecks, impacting performance incongruently across IP versions. Differences in performance between IP versions may not be due to the protocol itself, but rather, the infrastructure associated with the protocol and its measurement. However, prior speed test measurement work has not differentiated between the IP versions of measurements, and, thus, has not demonstrated if such variation exists [32, 42, 53, 54, 56, 59].

Furthermore, anecdotal evidence indicates that households experience differences across IPv4 and IPv6, with reports of each providing faster speeds in different situations [37, 45, 46]. Existing work that compares IPv4 and IPv6 performance focuses on latency [8, 24, 26, 38]. Although measuring network delay is a valuable baseline, throughput measures network capacity and presents another important dimension for understanding performance. However, measuring network speed presents unique challenges due to a lack of availability on client measurement platforms [49] and is more prone to variation based on factors such as subscription tiers [42]. As broadband deployment increases, comparing speeds across IP versions is important to evaluate Internet quality; results have the potential to impact policy decisions that help to narrow the digital divide [14].

We ask how systematic speed test measurements in home Internet connections differ across IP versions. We achieve this by performing repeated systematic IP-version-controlled Ookla [39] and NDT7 [34] speed tests from dual-stack households across multiple US states, ISPs, and CPE using the FLOTO Platform [25]. First, we characterize broad relationships between IPv4 and IPv6 speeds by analyzing which IP version is faster under different situations. Next, we identify the conditions under which IPv4 and IPv6 differences are more likely to occur, including the ISP and speed tier.

Findings. We find that the IP version used in a speed test affects its result. We identify that this impact is correlated with several factors, such as speed test software and the ISP. Although our measurements do not represent a generalized view of the Internet at large, they provide evidence that IP version reliably affects speed test results, and motivate a need to consider IP version as a primary factor in future explorations of home Internet performance. More specifically, from our measurement vantage, we identify several conditions where IPv4 and IPv6 speed results are similar or different. We highlight these trends in our results to illustrate *systematic* differences in IPv4 and IPv6 throughput.

Similarities. We find that IPv4 and IPv6 speeds differ by low margins in 81.7-91.5% of speed tests, depending on speed test parameters (software and the target server). We summarize these trends across the different facets we measure from our vantage points:

- We characterize a *steady-state* behavior for speed comparisons between IP versions. Across all data, IPv4 is most likely to be slightly (up to 2.5%) faster than IPv6. This slight difference is likely due to minor optimizations present in IPv4 that do not exist in IPv6.

- Because this steady-state behavior is relatively constant, we find that the consistency and individual household speed range are not affected by the IP version of the test. We find rare exceptions in high speed tiers (>600 Mbps in download speeds, >400 Mbps in upload speeds).

Differences. We find that speed tests differ significantly in a non-negligible fraction of cases (8.5–18.3%), disproportionally influenced by several factors, including speed test software and ISP.

- We find systematic differences in testing infrastructure. NDT7 tests are more likely to be faster in IPv4, while we found one Ookla speed test server that is more likely to report low speeds in IPv4. Some of these differences have a substantial impact on the interpretation of the results.
- The IP version of the speed tests has a significant effect on the overall speed test distribution in almost all households. Although significant differences (>5%) are less common, we demonstrate cases where they occur in conditions specific to the ISP. This includes specific cases that indicate a household's speed tier may play a role in which IP version is faster.
- We find that individual households are occasionally more prone to be significantly faster in one IP version. Although this is the minority of cases, this prompts us to hypothesize that local household-specific factors (such as local network conditions) affect which IP version is faster.

We additionally release our speed test data to encourage future work in understanding the role of IP version in measured Internet throughput [62]. We redact data that may compromise user privacy, such as full client IP addresses.

Our work calls for revisiting the impact of IP protocol on Internet speeds. We show that future research on broadband quality should consider IP version as a factor that can affect Internet performance. From these findings, we make several recommendations for future work measuring Internet speed using speed test software. As these measurements have consequences, such as when used to ground policies [11, 14], we emphasize the importance of understanding IP version in conjunction with coverage and quality. We hope that future work can leverage our findings to make Internet quality analysis more accurate and impactful.

2 Background and Related Work

To the best of our knowledge, despite numerous works analyzing Internet speed test methods and data, no prior work compared IPv4 and IPv6 Internet throughputs in residential networks with active measurements. We begin with an overview of the differences between IPv4 and IPv6. Next, we discuss broadband measurement, including speed test software, analysis, and best practices.

2.1 IPv4 vs. IPv6

In addition to differences in address size, IPv4 and IPv6 differ in several ways. Here, we highlight those that are relevant to this work, including network structure, use in applications, and performance comparisons in prior work.

Differences in IPv4 and IPv6 Networks. Despite IPv6 representing nearly 50% of client connections by Google [17], IPv4 and IPv6 network infrastructure differ significantly. Due to IPv4 exhaustion, network address translation (NAT) technologies, including Carrier-Grade NAT (CGNs), are deployed to extend the use of IPv4 addresses [48]. However, due to the ubiquity of IPv6 addresses, NAT is not necessary in IPv6 networks [57]. Because the Internet is now composed of both IPv4 and IPv6 networks, transition mechanisms are deployed to bridge the two IP versions [23]. Moreover, middleboxes are more likely to occur in IPv4 paths compared to those in IPv6 [21]. AS-level topologies may also differ [8]. Finally, a broadband network's IPv6 support may vary [44]. These differences provide us with the basis to suspect different bottlenecks exist in each IP version and motivation to compare throughputs.

Choosing IP Version in Dual-Stack Applications. In many dual-stack applications, a Happy Eyeballs algorithm is used to determine whether an IPv4 or IPv6 connection should be used [52, 64]. At a high level, this algorithm selects which IP version will be used by attempting connections over both versions and using the connection that is established first. IPv6 is prioritized by starting the IPv6 request first and allowing for a timeout period where, if the IPv4 response is received first, IPv6 will still be chosen. Prior work found that IPv6 is always preferred by client applications [2, 51] and highlighted the effect of DNS caching [22]. By comparing throughputs instead of latency, our findings add important context to IP version and performance.

Comparing IPv4 and IPv6 Performance. Several prior studies compared network latency by IP version from commercial or academic vantages. Prior work has measured latency to web servers [8, 16, 26, 38] and compared performance when transition mechanisms are used [63, 65]. However, many of these works are 5-20 years old, when native IPv6 adoption was much lower [17]. Some work included throughput but only used communication between two machines in a controlled laboratory setting [7, 36] or one measurement vantage towards dual-stack websites [27]. More recently, Huston explored IPv6 performance [24], but this study only includes connection success rate and RTT. In routing hardware, Cisco reports that IPv4 and IPv6 throughput are similar in larger packets, but IPv6 throughput suffers in smaller packets [5]. Contrary to these studies, we measure throughput from residential access networks.

2.2 Broadband Measurement

In this section, we review related work on broadband measurement. Prior work on Internet throughput and speed test datasets has been characterized and used to inform future measurements through recommended best practices.

Speed Test Software. Macmillan et al. describe the differences in the test design of the two speed tests we use, NDT7 and Ookla [30, 34, 39]. Most notably, to attempt to flood the network path, NDT7 uses a single TCP connection and runs for ten seconds, while Ookla uses multiple TCP connections and varies in time. Additionally, the servers used in NDT7 tests are managed by M-lab, while Ookla servers can be operated by any network. Finally, in both tests, the amount of data transferred in the TCP payload determines the reported throughput. However, NDT7 calculates this based on the entire duration of the test, and Ookla implements a sampling method to discard low throughput values, accounting for the TCP slow-start at the beginning of the test. Macmillan et al. further found that NDT7 tests are more likely to under-report speeds compared to Ookla, but that Ookla servers can potentially bottleneck results. We observe both of these effects and reflect on them in Sections 4 and 6.

Both NDT7 and Ookla can run in both IP versions. Between January 1st and October 2nd, 2024, 67% of NDT7 tests were IPv4 and 33% were IPv6. Publicly available Ookla data does not include the IP version of each test [40], preventing its users from distinguishing between tests.

Broadband Measurement and Analysis. Prior work has used speed test datasets to build maps to understand geographic differences in service. The FCC's Measuring Broadband America (MBA) program measured Internet speeds from selected volunteers located in the United States and reported on their findings yearly [14]. Fries et al. built a map of historical broadband performance over 8 years using data from the MBA project [15]. They identify several geographic areas that are underserved and make recommendations to improve connectivity. Other work queried ISP tools to understand broadband service offered across the US [31, 32, 41].

Several works have also found that various factors, including geographic location, CPE hardware, and ISP, impact residential Internet performance [9, 10, 53, 56, 59]. Other work has explored best practices for understanding data quality, including analysis methods for speed test datasets.

Sharma et al. show that *comparative* metrics between different geographic areas can be more valuable than simply calculating whether certain networks reach desired benchmarks [54]. Other best practices for understanding results include factoring in the latency, client type, and whether wireless or wired access was used [29], as well as using the speed tier to add context to the analysis [42]. We use several of these metrics in our analysis to understand IPv4 and IPv6 comparisons, as well as context on service plans offered to group our households into speed ranges.

3 Methodology

We now describe our measurement and data collection methods, beginning with an overview of the residential platform and a detailed description of our measurements.

3.1 Measurement Infrastructure

To execute speed tests in households, we use the FLOTO platform [25]. The FLOTO project recruits volunteers who allow network measurements to be conducted within their households. The measurements are executed on Raspberry Pis connected to the home routers via Ethernet. We deploy our measurement software to the devices via Docker containers using the FLOTO platform. At the time of our measurements, the majority of FLOTO volunteers were located in two major US cities across a diverse set of neighborhoods.

The FLOTO platform allows us a unique measurement vantage inside households directly. Although other platforms have existed that measure clients, these platforms are either not currently operational [12, 58] or do not allow throughput measurements [49]. Moreover, publicly available speed test data does not specify the IP version of the speed test [40] or does not contain speed tests in both IP versions from the same client that can be compared (i.e., executed closely in time) [13, 33, 40]. The FLOTO platform allows us to overcome limitations presented in existing data or other client measurement platforms to compare IPv4 and IPv6 speeds directly. In this work, we only use measurements from residences with both IPv4 and IPv6 connectivity (i.e., dual-stack networks). This leaves us with data from 60 households in total.

3.2 Ethics

The FLOTO platform was created with Institutional Review Board (IRB) approval and the consent of the participants. The data collected by FLOTO is anonymized and only includes active measurements. We collect no data about the other network traffic in the residence, and our measurements were done within the parameters of the platform consent agreement. Volunteers are compensated financially, given access to a private web application that they can use to monitor their own home Internet, can withdraw from participation at any time, and are given a point of contact at the FLOTO project.

3.3 Measurements: Paired Speed Tests

We use two speed tests: NDT7 [34] and Ookla [39]. As discussed in Section 2.2, these tests differ in their design and, consequently, their results. Based on prior observations that Ookla speed test server has an effect on measurements [30], we experiment with different speed test servers from three ISPs: Comcast, Charter, and AT&T. We choose these speed test servers because they are congruent with the broadband providers for the majority of our households. We distinguish tests from these three different servers by referring to them as "Comcast Ookla," "Charter Ookla," and "AT&T Ookla". As NDT7 doesn't allow one to specify an IP type, we modify NDT7 to allow us to specify a source address, enabling testing in both IPv4 and IPv6. Ookla speed tests allow for specifying an IP to bind to; however, we encounter configuration errors with this option and instead specify the server IP in each device's host configuration file.

To capture a variety of network conditions, we execute our speed test measurements over one week and execute 3 paired speed tests of each type (NDT7 and Ookla using the 3 different servers) per hour. These paired speed tests are IPv4 and IPv6 tests conducted consecutively, with a brief cool-down period in between. This way, our results are directly comparable but do not interfere with each other. Tests are executed in time at a unique offset for each household to ensure that the speed test servers are not overwhelmed or bottlenecked by our tests. That is, the speed tests are conducted at regular intervals, but start at different times.

For Ookla tests, we always run IPv6 tests before IPv4. For NDT7 tests, we keep the order constant, but that order differs per device. i.e., one device might run IPv4 then IPv6, another device might run IPv6 then IPv4. We repeat our experiment at a smaller scale at a slower pace (1 paired speed test every 3 hours) with random test orders and find no difference correlated to the order of the test. We attribute this to the cool-down period.

3.4 Metrics to Understand Speed

To understand the differences in speeds observed from our various speed tests, we use a set of metrics derived from prior work that analyzes measurement performance [30, 42].

Relative Speed. Given our paired speed tests, we directly compare throughput in IPv4 and IPv6 by calculating the relative speed per pair. Following prior work [30] that analyzed relative differences between speed test software, we calculate relative speed, δ, by:

$$\delta = \frac{\text{IPv6 Speed} - \text{IPv4 speed}}{maximum(\text{IPv4 speed}, \text{IPv6 speed})}$$

Using this, positive values (>0) indicate IPv6 is faster, and negative values (<0) indicate IPv4 is faster. For example, if relative speed is -0.10, IPv4 is 10% faster.

Normalized Speed. To understand how individual speed test results change with respect to the expected performance per household, we calculate the normalized speeds. Once again, following prior work [30], we calculate normalized speed by dividing the measured speed by the 95th percentile speed for the IP version in an individual household. This metric intuitively represents how a speed test compares to the expected performance for a household.

Normalized speed and relative speed do not necessarily coincide. For example, IPv4 normalized speed could be 0.8, but relative speed could still be -0.1, indicating IPv4 is 10% faster than IPv6 while still slower than we expect.

Consistency Factor. To understand how consistently an individual household reports a speed, we calculate the consistency factor. As calculated by prior work [42], we calculate the consistency factor with the mean speed of all tests divided by the 95th percentile speed over all tests. With this metric, a high value indicates consistent speed reporting, while a low value indicates the opposite.

3.5 FLOTO Platform Calibration

A natural question when using a platform such as FLOTO to measure speed is whether the platform itself has an effect on our observed speeds. In this section, we describe the results of an experiment to evaluate the FLOTO platform itself.

3.5.1 Methodology. We measure and compare speeds with a FLOTO-connected Raspberry Pi and a non-FLOTO Raspberry Pi (which we refer to as the “control”) in the same network. To mirror the conditions of the FLOTO devices in our experiments, we perform these measurements in a residential environment in a major US city. We mimic the conditions of our large-scale speed test measurements by executing speed tests using NDT7 and Ookla using a speed test server in the same ASN as the household service. We use the same speed test server for the tests from the FLOTO device and the control device. We perform paired speed tests on these devices in IPv4 and IPv6. We run both speed tests within a few minutes of each other to capture the same network conditions and compare the results. On each device, we execute 3 paired speed tests of NDT7 and Ookla per hour for approximately 48 hours. In total, we collected 140 paired speed tests to compare across FLOTO and control.

As our goal is to understand the differences between IPv4 and IPv6 speeds, we evaluate whether there is a statistically significant difference in the relative speeds recorded by the FLOTO platform and the control. Similar to prior work that compares NDT7 and Ookla speed tests [30], we use a paired t-test. We satisfy the assumptions that the tests are conducted in pairs and we have at least 30 tests, satisfying the normality condition [43]. The null hypothesis is that there is no difference between the mean relative speeds between the FLOTO platform and the control. We use a significance level of 0.05, meaning that we reject the null hypothesis if our p-value is less than this threshold [43].

We further calculate the Cohen's d effect size [6]. Effect size values under 0.2 are considered very low, 0.5 is considered moderate, and over 0.8 is considered large [6]. This allows us to quantify any differences, even if they are not statistically significant.

3.5.2 Results. In every case, we fail to reject the null hypothesis. This includes both NDT7 and Ookla speed tests for download and upload relative speeds. We also observe low effect sizes in all cases. In all testing parameters (upload, download, and each software), the effect size is 0.05 or less. The only exception is Ookla upload tests, where the effect size is 0.22, which is still negligible [6].

We find that the FLOTO platform does not impact the distribution of the relative speeds, and therefore also does not impact our desired speed test measurements. Thus, we can leverage this platform for our investigation.

3.6 Limitations

Our vantage is subject to limitations by geographic location (two major US cities) and the number of households within three ISPs. However, we highlight that this is a common number of vantage points for similar work in residential speed tests [30, 53, 54], and performance [19, 50, 55, 60] analysis and measurement, especially those that require controlled, regularly executed experiments such as ours. As stated in Section 3.1, we again emphasize that few other platforms exist that allow for the speed tests we execute. While measuring latency can be done from community-driven platforms such as RIPE Atlas to a variety of different destinations, speed tests are not included in the pre-defined set of measurements that are allowed [49]. Other platforms that allow a wider set of active network measurements are limited to the networks of universities and have as few as 50 vantage points [3]. Thus, the size of FLOTO is comparable to the few alternatives that exist. Furthermore, in all volunteer platforms, there is a natural selection bias. That is, the population that volunteers to participate in these measurements is not necessarily representative of the overall population of Internet users.

We additionally acknowledge the effect that speed test configurations may have on our results. Prior work has shown that factors such as ISP, router vendor, speed test server, and time period affect the test result [20]. Particularly, prior work comparing speed test results has found that Ookla speed test server choice can result in lower performing speed tests [30]. Prior work in residential Internet performance does not specify the exact ISP distribution, but use a broader set of speed test servers [30], have a wider geographic sampling [19, 60], measure a single router vendor [60], or analyze different applications [50]. Some work [19] discusses the geographic sampling in more detail or has wider samples. RIPE Atlas has a diverse set of over 12,000 globally distributed probes in geographic areas across many different providers, but does not allow for speed test measurements [49]. Other platforms that allow these types of measurements also do not publicize data such as the ISP, although they do reveal geographic information and have the same device type as our measurement platform [3].

Although we lack comprehensive data across the Internet at large, our vantage allows us to reliably explore IP version as a factor that impacts speed test measurement reliably. Thus, while the specific trends we identify and measure may not generalize to other networks, our broad result that IP version impacts Internet speed still holds.

Due to deployment constraints, we cannot capture all possible network conditions. We also note that we are limited to the dual-stack households that are a part of the FLOTO platform. Moreover, in the case of RIPE Atlas as well as the FLOTO platform, special-purpose measurement devices are not necessarily representative of what clients use. Because of these constraints, rather than claim the trends we observe are congruent with the wider Internet, we use our findings to identify factors that *can* affect a measurement result.

While our measurements are from two cities, the households that we measure from cover a diversity of speed tiers, neighborhoods, and router vendors. We do not believe that the number of cities that we measure from is a limitation, and we do not expect that different cities would introduce different effects. Furthermore, we trade off a higher number of households in favor of ensuring representation by router vendors and providers. This allows us to evaluate the effect, or lack thereof, of these two factors on differences in speed by IP version.

In summary, the scale of our vantage points provides enough data for meaningful longitudinal analysis representative of multiple providers and routers. This work provides an initial investigation and serves as a basis for future work to expand measurements to a wider variety of environments.

4 Data

In this section, we describe our speed test dataset collected using FLOTO. We begin by describing the demographics of the households conducting our measurements. Next, we provide an overview of our speed test results, including categorizing the households into speed tier ranges and analyzing the consistency of the reported speeds per device.

4.1 Household Demographics

We now discuss the characteristics of the households that participate in our measurements. To avoid using one-off cases to generalize our findings, we ensure that our dataset has a representative number of households per broadband provider and router vendor. In total, this leaves us with 87,173 paired speed tests from 46 households.

Broadband Provider. We only consider providers where we have at least 10 households participating in our measurements. We have 21 in AT&T (ASN 7018), 14 in Comcast (ASN 7922), and 11 in Charter Communications (ASN 10796).

Identifying Router Vendor. We also filter for CPE vendors to avoid outliers from specific hardware implementations. We only include data from households where we have at least 5 others with the same router vendor. We use Zgrab [61] to make HTTP and HTTPS requests from the devices toward their router's private IP in IPv4 and IPv6, identified through traceroutes. We manually analyze the HTTP body and TLS certificates to identify the vendor. We have 17 households with an Arris router (the manufacturer for the AT&T routers [1]), 14 that use an Xfinity router (used with Comcast Internet service), 5 with a Sagemcom router, 5 with an ASUS router, and 5 with another AT&T-given router (distinct from the Arris routers identified from their certificates).

Geography. All of the households are located in one of two major US cities which we refer to as City A and City B. All of our data in ASN 10796 (Charter) are in City A, and all of our data in ASN

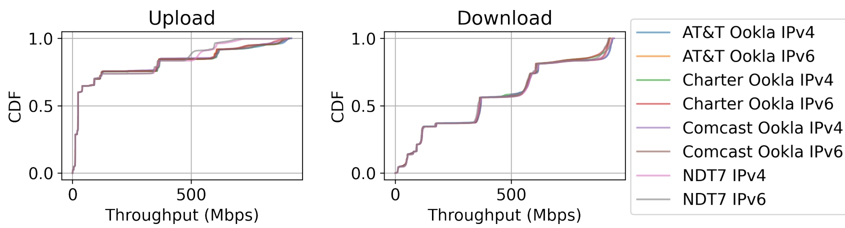


Fig. 1. CDF of all throughputs. We observe several common download and upload speeds. In most cases, we see no aggregate differences between IP versions and testing parameters. One exception is NDT7 upload speeds, where we observe fewer speeds above 443 Mbps.

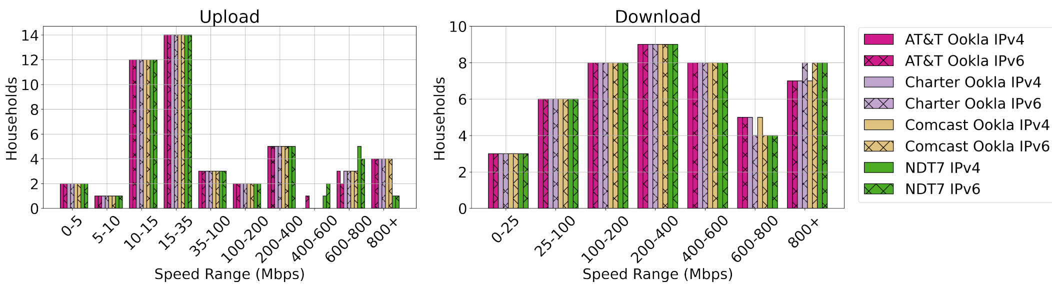


Fig. 2. The speed range for each household, calculated using the 95th percentile speed. We observe general consistency for the speed range between testing parameters. We find exceptions only in high speeds above 600 Mbps download and 400 Mbps upload.

7922 (Comcast) are in City B. In ASN 7018 (ATT), 12 households are located in City A and 9 are located in City B. We note that our vantages cover a variety of neighborhoods within these cities.

4.2 Speed Tests

After running our speed test measurements 3 times per hour for a week in April 2024, we have an average of 460 paired speed tests per type (NDT7, Ookla with 3 different speed test servers) per household, with a standard deviation of 94. This number varies because of temporary measurement gaps, e.g., due to a user accidentally unplugging the device.

Figure 1 shows an overview of all the throughputs. We observe several modes at different speeds. In download speeds, we observe no high-level differences between testing parameters. However, in upload speeds, we see that NDT7 has fewer upload speeds past 443 Mbps compared to Ookla.

Speed Tier Ranges. We next categorize households into ranges of speeds and discuss the effect of IPv4 or IPv6, as well as testing parameters. We use the 95th percentile of the observed speeds from each household to determine the general expected speed. Given our regular measurements, we use this as a proxy for calculating the speed tier (i.e., the subscription tier) and categorize households in speed ranges. We choose ranges based on prior work [42] that queries ISP subscriptions and add higher tiers to incorporate the higher speeds we observe in our households. We find that the general range of speed for the household does not change depending on the speed test type (NDT7 or Ookla) or speed test server in Ookla measurements, although we identify a few abnormal cases.

We show the overall distribution of ranges in Figure 2. Although a few households have discrepancies in categorization depending on testing parameters, we find that these are not a result of "borderline" cases. We also note that, given the high number of speed tests per household, this is

			Med.	>0.9	Med. Δ	Δ <0.01
AT&T Ookla	↓	IPv4	0.97	86.9%	0.0004	69.6%
		IPv6	0.98	91.3%		
	↑	IPv4	0.98	91.3%	0.0008	93.5%
		IPv6	0.98	91.3%		
Charter Ookla	↓	IPv4	0.97	84.8%	0.001	86.9%
		IPv6	0.97	86.9%		
	↑	IPv4	0.98	91.3%	0.0008	93.5%
		IPv6	0.98	89.1%		
Comcast Ookla	↓	IPv4	0.98	93.5%	0.001	82.6%
		IPv6	0.98	93.5%		
	↑	IPv4	0.98	89.1%	0.001	89.1%
		IPv6	0.98	89.1%		
NDT7	↓	IPv4	0.97	95.6%	0.002	95.6%
		IPv6	0.97	95.6%		
	↑	IPv4	0.97	86.9%	0.002	91.3%
		IPv6	0.98	84.8%		

Table 1. Here, we show various statistics about the consistency factors (defined in Section 3.4) for each set of speed test parameters. For brevity, we denote upload speeds as ↑ and download speeds as ↓. The "Med." column shows the median consistency factor, ">0.9" shows the percentage of households with a consistency factor above 0.9, the "Med. Δ" column is the median difference in consistency factors in IPv4 and IPv6, and the "Δ <0.01" column shows the percentage of households with a difference under 0.01. We observe that, generally, median consistencies are high and do not differ significantly between IPv4 and IPv6.

not the result of a few low throughput results. We also find that these cases exclusively happen at higher speeds above 600 Mbps in download and above 400 Mbps in upload.

In download speeds, only one household has a discrepancy between test parameters. This household is in ASN 7922 (Comcast) and tests significantly lower (654–740 Mbps) in Comcast Ookla IPv4, AT&T Ookla in both IP versions, and Charter Ookla in IPv4 compared to others (857–902 Mbps). In upload speeds, 5 households vary in speed range depending on testing parameters. All of these are in ASN 7018 (AT&T). On average, the difference between the highest and lowest expected speed for these households is 172.8 Mbps across all testing parameters. Aligning with prior work [30], we note that NDT7 upload speeds are more likely to be lower than Ookla tests.

As such cases are a minority (13% total, 1 in download and 5 in upload speeds), we determine that, overall, classification remains largely consistent across test parameters.

Speed Consistency. To understand the variation of speed test results from an individual household, we calculate the consistency factor for each combination of speed test, test type, upload or download, and IP version.

We plot various statistics about the consistency factors in Table 1. We observe that the vast majority of consistency factors calculated in every test situation (speed test, server, upload or download, IPv4 or IPv6) are high, and all median values are 0.97 or higher. In every situation, we also observe that at least 84% of households have a factor above 0.9. Therefore, results outside of expected performances are rare. We additionally calculate the difference between the IPv4 and IPv6 consistency factors in each case. In most cases, we observe small differences (<0.1) between calculated factors. We highlight that download tests in AT&T Ookla are more different than other types of tests, but still differ by 0.1 or less in almost 70% of cases.

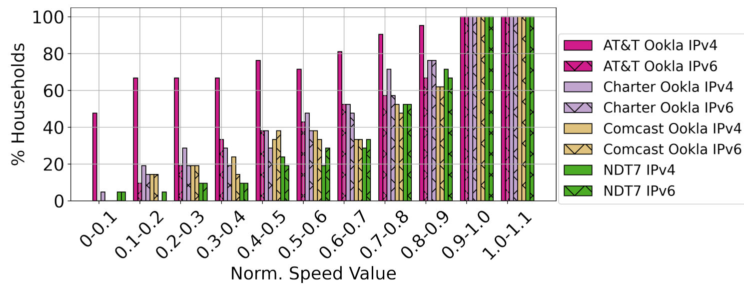


Fig. 3. Percent of AT&T (ASN 7018) households with at least one download speed test in each normalized speed bin. These households are more likely to have low values in IPv4 using the AT&T Ookla server.

We investigate this further by analyzing the normalized speeds for speed tests conducted using this server. We show this distribution for households in ASN 7018 (AT&T) in Figure 3. We observe that 100% of AT&T households have at least one speed test with an IPv4 normalized speed under 0.6. When using other testing parameters, at most 52.4% of AT&T households have an IPv4 normalized speed under 0.6. We therefore determine that AT&T households are more likely to observe lower speeds when using the Ookla AT&T server. We emphasize that this pattern only appears for AT&T households using the AT&T Ookla speed test server. When using other speed test parameters, only 34.7–47.8% of AT&T households have any normalized IPv4 download speed test under 0.6. Moreover, only 43.7% of households in ASN 10796 and only 44.4% of households in ASN 7922 have any IPv4 download speed test with this value when using the AT&T Ookla server. We repeat this experiment and execute speed tests at a much lower rate (1 paired speed test every 3 hours) with a longer cool-down time between tests, and a random test order. We additionally use a different AT&T speed test server that is within the same city. We observe the same result where AT&T tests are more likely to perform worse in IPv4 tests, and therefore conclude that this effect does not emerge from overwhelming a single speed test server in our testing methodology. Although we cannot identify the cause, we therefore hypothesize that there may be an IPv4-only path bottleneck within AT&T's network.

While the consistency factor does not directly quantify the differences in speed (i.e., which is faster), this indicates that IPv4 and IPv6 are not inherently more or less consistent than one another. That is, even if one IP version is faster, it is consistently faster, and the other is consistently slower.

5 Results

In this section, we discuss our analysis of paired IPv4 and IPv6 speed tests. We first present a high-level overview of the effect of IP version in our overall dataset and how paired speed tests compare. We also highlight patterns in speed test performance per household. We then discuss *consequential* differences in speeds recorded in different IP versions and characterize the conditions where these differences occur. We finally discuss other characteristics measured alongside the paired speed tests to understand their relationships to differences in throughput.

While we find that in most cases IPv4 and IPv6 speeds are comparable, we find that some households more frequently exhibit speeds faster in one IP version, and IPv4 and IPv6 performances diverge in patterns specific to providers. We also present a taxonomy for understanding differences in speed performance and leverage it to help us understand speeds in different conditions.

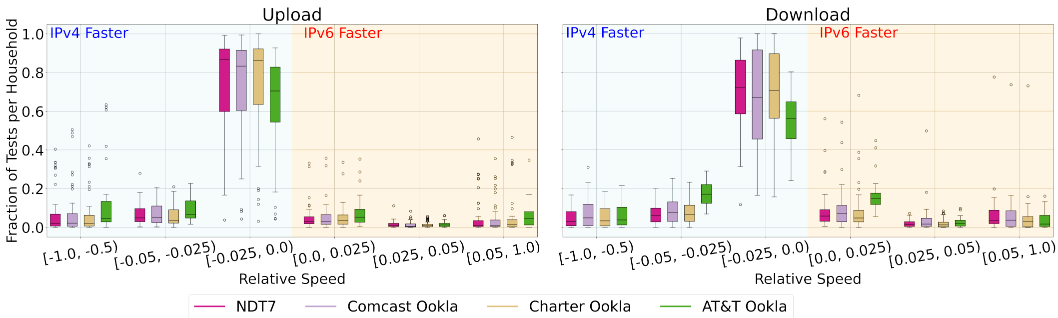


Fig. 4. Fraction of paired speed tests that fall within the relative speed range for a single household. Values < 0 are faster in IPv4, values > 0 are faster in IPv6. We observe that the bulk of our tests are slightly faster in IPv4. We highlight that a relative speed value of 0 would be on the right side of the plot (in the bucket `[0.0, 0.025)`); however, we have no such values.

5.1 Differences in IPv4 and IPv6

We begin by comparing the overall dataset of paired IPv4 and IPv6 speeds. We first analyze the relative speed values (defined in Section 3.4) to determine how pairs of IPv4 and IPv6 speeds compare. Next, we show the statistical significance of the IP version on the overall distribution of speed test results per device. Finally, we discuss the relationship between the relative speeds of paired speed tests and the normalized speeds of the speeds in each IP version, finding that *speeds in each IP version degrade under different conditions.*

Relative Speed. To compare IPv4 and IPv6, we calculate the relative speed for each paired speed test. That is, for each paired speed test, which is faster, and by how much?

For each type of test, we calculate the fraction of tests per household in different buckets of relative speed values. Figure 4 shows our results. In each case, the most common relative speed is less than 0, but greater than or equal to -0.025. This indicates that IPv4 speeds are usually 0–2.5% faster than IPv6. Although we cannot determine the exact cause of this, we note that we observed the same effect when evaluating the effect size of the FLOTO platform in Section 3.5, and in small-scale experiments on different systems (i.e., on laptops running Windows and macOS). This discrepancy may result from protocol differences, such as the larger header size in IPv6 packets (40 bytes), as opposed to 20-60 bytes in IPv4. However, we also emphasize that a 0–2.5% difference in measured speed is, in most cases, 2–10 Mbps based on the speed ranges of our households in Figure 2. Although in these cases the effect of this difference is limited, at higher speeds, this difference is more likely to noticeably decrease Internet quality.

IP Version Impact on Test Distribution. Next, we compare the full distribution of speed tests in each IP version per household. We determine whether there is a statistical significance using a paired t-test and calculate the effect size. Thus, we ask: 1. Is the difference between IPv4 and IPv6 speeds statistically significant (significance)? 2. How different are these speeds (effect size)?

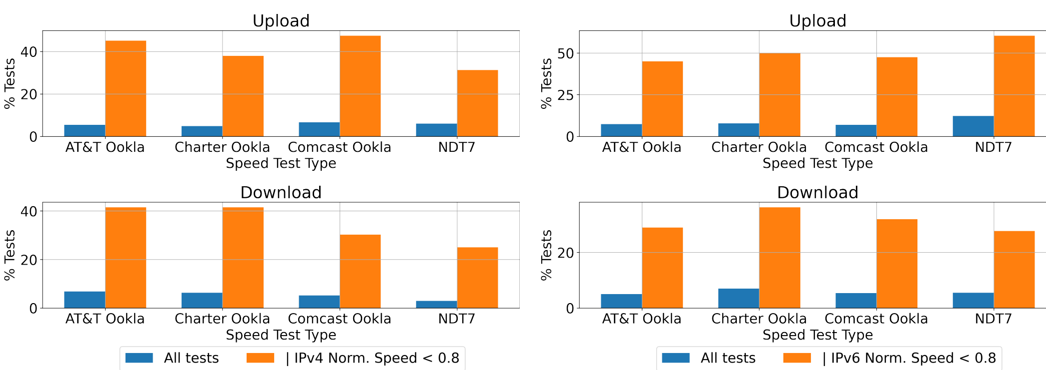
As in Section 3.5, we begin by using a paired t-test to compare the IPv4 and IPv6 speeds recorded per household. Once again, the null hypothesis is that the speed distributions do not differ. We show the percentage of households that rejected the null hypothesis in Table 2. We show that the majority of households have statistically significant differences in their IPv4 and IPv6 speeds. Next, we calculate the effect size for the IP version of the speed test for each household. In Table 3, we show the percentage of households where the effect size is over 0.2, indicating the effect is sizeable [6]. This observation applies to most households in most of the testing situations.

	AT&T Ookla	Charter Ookla	Comcast Ookla	NDT7
Download	47.8%	76.0%	71.7%	80.4%
Upload	82.6%	80.4%	86.9%	71.7%

Table 2. The fraction of households that reject the null hypothesis in a paired t-test between versions' speeds. We find that, generally, IP versions yield different speeds.

	AT&T Ookla	Charter Ookla	Comcast Ookla	NDT7
Download	32.6%	58.7%	60.9%	71.7%
Upload	71.7%	69.6%	67.4%	65.2%

Table 3. The fraction of devices that have an effect size > 0.2 when comparing IPv4 vs IPv6 throughput.



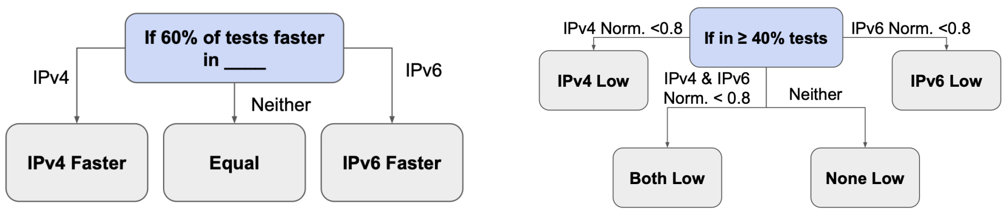
(a) Fraction of tests where IPv6 5%+ Faster (b) Fraction of tests where IPv4 5%+ Faster

Fig. 5. Percent of paired tests with differences > 5%, and percent of paired tests with differences > 5% when one IP version's normalized speed is < 0.8. There is a significant increase in paired tests with a high relative difference when one IP version is slower than expected.

Relating Relative Speed and Normalized Speed. With the understanding that IP version has a statistically significant effect on speed, we seek to understand these differences with respect to the overall speed. That is, when speed in one IP version degrades, does the relative speed change?

To answer these questions, we compare the relative speeds to the normalized speeds. For example, if normalized speed shows that IPv4 is performing at 80% of the expected speed, a relative speed value of 0.5 will indicate that the IPv6 speed is much faster and likely has not also decreased. In Figure 5, we illustrate these comparisons when IPv4 or IPv6 is at least 5% faster. We observe that, while relatively few paired speed tests differ by over 5% in favor of a specific IP version (i.e., 7% or less in all cases have a relative speed greater than 0.05, *and* 7% or less are less than -0.05), that fraction significantly increases to 25% or more in all cases when we filter for normalized speeds below 0.8 (80% of expected speed) in IPv4 and IPv6.

We emphasize that this behavior indicates a divergence in performance between IP versions. That is, **IPv4 and IPv6 performance is decreasing in different instances**. We hypothesize that this could be in the local network (i.e., congestion in a queue for one IP version and not the other), a path bottleneck, or congestion at the speed test server.



(a) Thresholds for labeling the faster IP version

(b) Thresholds for labeling the normalized speed

Fig. 6. An overview of the taxonomy we use to categorize groups of speed tests. Based on our findings that speeds degrade under different conditions, we label groups according to two factors: which IP version is faster and whether there are low normalized speeds.

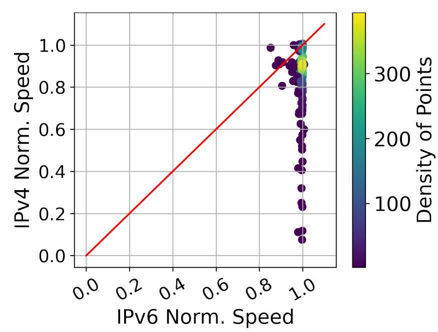


Fig. 7. Normalized speeds from paired Comcast Ookla tests from a single household. The red line represents where the normalized speed in both IP versions is the same. Values above the red line are faster in IPv4, and values below are faster in IPv6.

Takeaway: There is a *steady-state* behavior that most households follow where IPv4 speeds are up to 2.5% faster than those in IPv6 (Figure 4). However, when speeds deteriorate, they do not do so in conjunction across IP versions (Figure 5).

5.2 Taxonomizing Differences in Speed

Following our understanding of the relationship between relative speed and normalized speed, we seek to characterize how these two characteristics compare for different groups of paired speed tests. To systematically label these changes, we define a taxonomy in order to compare how different factors change overall behavior in our subsequent results.

Figure 6 shows an overview of our taxonomy. We characterize which IP version is faster and whether either IP versions have a low normalized speed. To characterize which IP version is faster, we use a threshold of 60% of tests. That is, if 60% or more tests are faster in IPvX, we label it as faster. If neither IPv4 nor IPv6 is faster in at least 60% of tests, we label the speeds as equal. To characterize low normalized speeds, we similarly use a threshold of 40%. If 40% or more tests have a normalized speed under 0.8, we consider the group to have a low normalized speed in that IP version. We choose 0.8 as our threshold because 80% of the 95th percentile of all its speeds is significantly lower and likely impacts user experience. We choose the thresholds based on modal shifts in the distribution; Appendix A expands this method.

Motivating Example. To illustrate why we chose these parameters, we analyze the paired speed tests from one household in Figure 7. Here, we show the distribution of normalized speeds in IPv4 and IPv6. For brevity, we show only the results from Comcast Ookla speed tests. We note that we do not plot the relative speed, but, in this case, it coincides with relative speed where 87.3% of relative speed values are faster in IPv6 and 88.7% of normalized speed values are higher in IPv6 than IPv4. We observe that the most concentrated area in this graph is tests that are faster in IPv6 (87.3% of tests), however, there are outliers that are faster in IPv4. We additionally observe many tests that have low normalized speeds in IPv4. However, these only cover 8.3% of all tests, all of which are below 0.8. Therefore, our characterization for this example would be no low normalized speeds, faster in IPv6. Although there are outliers that do not follow this characterization (i.e., 12.7% of tests are faster in IPv4, and 8.3% of tests have low normalized IPv4 speed), our labels capture the general behavior for this household.

To label speed performance in an individual IP version and relative speed, we define a taxonomy that we apply to different groups of speed tests in Sections 5.3 and 5.5.

5.3 Household-Specific Speed Effects

To understand an individual household's speed with respect to the IP version, we apply our taxonomy to the data from each household. This allows us to understand how general speed changes for households individually.

Following the steady-state conditions described in Section 5.1, we find that the majority of our households have faster download and upload speeds in IPv4 by our taxonomy (>60% of tests). In Ookla tests, 89% of households are faster in both download and upload speeds. In NDT7, 93.5% of households have faster download speeds in IPv4 and 95.6% have faster upload speeds.

We also find that most households are labeled with no low normalized speeds when analyzing all of their measured paired speed tests. In Ookla tests, this covers 93.5% of households in both download and upload. In NDT7 tests, 95.6% have no low normalized download speeds and 87.0% have no low normalized upload speeds.

Some households are outliers in both Ookla and NDT7 tests, i.e., they are not faster in IPv4 or have low normalized speeds. In Ookla tests, 4 households (8.6%) switch between IPv4 and IPv6 being faster depending on the server used in both download and upload measurements, and one household is faster in IPv6 download speeds. In NDT7 tests, the remaining households in both cases (2 in upload, and 3 in download) are equivalent in IPv4 and IPv6 speeds. In Ookla tests, 2 households have low normalized speeds in both IP versions, and one has low normalized IPv4 speeds in tests from the Charter Ookla server, and none using other Ookla servers. In NDT7 tests, the remaining households (2 in download, 6 in upload) had low normalized speeds in both IP versions.

From this analysis, we conclude that the majority of households follow our steady-state observations. However, we emphasize that the majority of these speed tests are marginally faster in IPv4 (<5%). We highlight that a low number of households have faster speeds in IPv6, making them outliers in our particular set of data. Additionally, a low number of households are more likely to have low normalized speeds, meaning they are slower than expected in more than 40% of tests. We use this as a demonstration that certain households are likely to have better Internet quality in IPv6 than others. We additionally do not see consistency in these occurrences across providers or CPE. We hypothesize that this is due to local network factors that we are unable to measure, e.g., a network bottleneck close to the household, or congestion within the home network itself due to extremely high traffic.

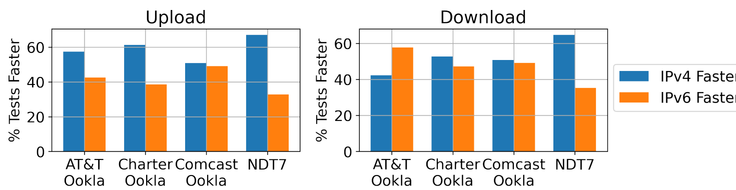


Fig. 8. The percentage of tests where each IP version is faster when |relative speed| > 5% in each set of speed test parameters. In most cases, IPv4 is faster. However, in NDT7 tests, the fraction of tests where IPv4 is faster (over 20%) is considerably more than in Ookla tests.

Takeaway: By taxonomizing the tests from each household, we see that most houses have faster speeds in IPv4 and speeds in either IP version do not broadly decrease in performance. However, a low number of outliers are prone to low normalized speeds and faster IPv6 speeds. Our findings indicate that, while most households share the same relationship in their IPv4 and IPv6 speeds, exceptions occur that indicate household-specific bottlenecks in a single IP version.

5.4 Considerable Differences by IP Version

With our understanding of the *steady state* relative speed behavior that is slightly faster in IPv4, and observations on IPv4 and IPv6 performance degrading differently, we now explore the conditions in which these differences occur.

Because we are most interested in significant differences in performance based on the IP version, we focus the following analysis on paired speed tests with a relative speed over 0.05 or under -0.05 (a 5% difference in speed). We choose this threshold to filter out the *steady-state* behavior and also to ensure that the tests we analyze have significant differences in speeds (i.e., where they have more potential to noticeably impact user experience). In Appendix A, we also highlight a modal shift in overall relative speeds at this threshold.

After such filtering, we have 10–13% of Ookla tests and 8–18% of NDT7 tests, depending on whether we analyze upload or download results. We expand on more differences we observe between NDT7 and Ookla tests in the following sections. We also highlight that, although this is the minority of all tests, this filtering leaves us ample data to analyze (1,815–3,911 *paired* speed tests, up to 7,822 total measurements). Table 4 in Appendix B provides more details of this.

We additionally acknowledge that some households tend to be represented more than others after filtering. Because some households rarely have speed test differences greater than 5%, they are either only in the filtered dataset once or, in some cases, not at all. However, these cases are uncommon, and most households are represented. In all of our filtered datasets, we retain tests from at least 41/46 (89%) of our households. Additionally, at least 28 households have at least 10 speed tests in each data set (although 34.8–53.6% have at least 30, satisfying the normality condition [43]). Thus, we consider the filtered version of the dataset representative.

The Faster IP Version When Differences Occur. After this filtering, we first observe the fraction of paired tests that are faster in IPv4 or IPv6. We show our results in Figure 8. Whether IPv4 or IPv6 is faster is not consistent per speed test parameters, and in most cases, IPv4 and IPv6 are equally likely to be faster in a given test. An exception is NDT7, where in both download and upload tests, over 60% of tests are faster in IPv4. We expand on differences in speed test software in Section 6.1.1.

We do, however, observe that more distinct patterns emerge when we separate results by the ASN of the household (i.e., the broadband provider) in Figure 9. In all Ookla tests from ASN 7922 (Comcast) households, IPv6 is more likely to be faster in download speeds, but in upload speeds,

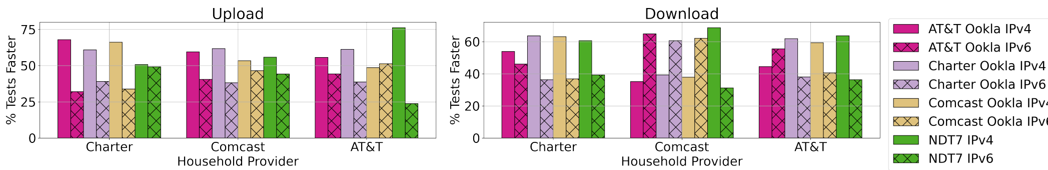


Fig. 9. The percentage of tests where each IP version is faster when |relative speed| > 5%, separated by the household's broadband provider.

IPv4 is more likely to be faster. In all Ookla tests from ASN 10796 (Charter) households, IPv4 is more likely to be faster in both download and upload speeds. However, in Ookla tests from ASN 7018 (AT&T) households, the faster speed changes depending on the speed test server. For example, the AT&T Ookla server is less likely to report faster speeds in IPv4 download tests than the Charter and Comcast Ookla servers. Finally, in all cases except for ASN 10796 and ASN 7922 upload speeds, NDT7 tests are faster in IPv4 in over 60% of tests.

Takeaway: Although large differences (>5%) between IPv4 and IPv6 are rare, when they do occur, the IP version that is faster varies per broadband provider (Figure 9).

5.5 Conditions for Speed Differences

We aim to understand what conditions lead to significant differences in IPv4 and IPv6 speeds. To do this, we analyze the paired speed test with differences over 5% (as described in Section 5.4) and filter by different factors, including the ISP and speed range. Then, we taxonomize the groups of results for comparison. This allows us to understand the conditions where *significant* divergences in speed occur and provides a high-level characterization. We find that patterns in our taxonomy labels emerge within each broadband provider.

ISPs. We first apply our taxonomy to the households separated by the broadband provider: Charter (ASN 10796), Comcast (ASN 7922), and AT&T (ASN 7018). Instead of simply presenting the percentage of tests that are faster in each IP version in Figure 9, our taxonomy labels represent differences and also account for normalized speeds. That is, our labels allow us to identify general shifts in speed test results. We find that taxonomized results are relatively consistent per ISP in all testing parameters with some exceptions. We first highlight the most dominant labels for the households in each provider for all testing parameters:

When speeds differ significantly, the majority of NDT7 tests are always faster in IPv4. In results from Ookla tests, the dominant labels are:
Charter (ASN 10796) & AT&T (ASN 7018): Download and upload speeds are either faster in IPv4 or are equal. Neither provider has low normalized speeds in either version.
Comcast (ASN 7922): Download speeds are faster in IPv6, upload speeds are equal.

In Charter and AT&T households, we describe the specificities of our labels with respect to the speed test parameters. In Charter (ASN 10796) households, we find two cases where both IP versions are equal: 1. when using the AT&T Ookla server for download speeds, and 2. NDT7 for upload speeds. In AT&T (ASN 7018) households, using the AT&T or Charter Ookla server elicits equal speeds, but using the Charter Ookla server or NDT7 elicits faster results in IPv4. We do not identify broad results with low normalized speeds with two exceptions: 1. with the AT&T Ookla server, where IPv4 is lower in download tests, and 2. with NDT7, where IPv6 is lower in upload

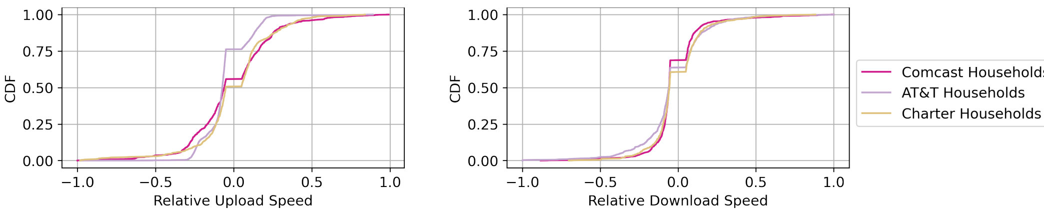


Fig. 10. CDFs of relative speed values for NDT7 tests per ASN of the household after filtering for paired tests with over 5% difference (thus the missing data from -0.05 to 0.05). We observe that the majority of our tests are faster in IPv4 (i.e., a positive relative speed value).

tests. In Comcast (ASN 7922) households, we find that an exception in upload speeds when using the Charter Ookla speed test server, where IPv4 is faster.

To illustrate the pattern in NDT7 tests that we find by using our taxonomy, we plot the relative speed results for each provider in Figure 10. We observe that in all download results, IPv4 is faster the majority of the time. In upload speeds, we observe that in Charter and Comcast households, both IP versions are equally likely to be faster or slower.

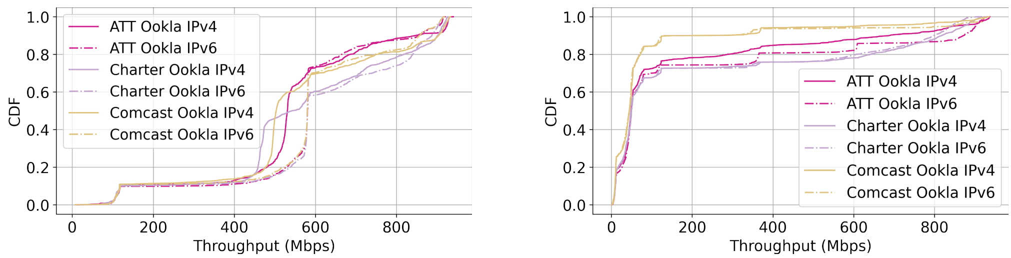
Because of the distinct patterns we observe that depend on the household provider, we thus focus the following analysis on each provider individually.

Speed Tier Ranges. We next use our taxonomy to examine how our results change when we separate our results by the speed range of the household. To satisfy the normality condition, we only analyze speed tier ranges that have at least 30 tests and 3 or 4 households. Since some providers have more households than others in our dataset, we change the number of households that we require based on the ISP. For households in Comcast (ASNs 7922) and AT&T (ASN 7018), we use 4 households, and in Charter (ASN 10796), we use 3. Per Figure 2, many of our speed ranges do not satisfy our condition for the number of households; thus, we aggregate to new ranges. For upload speeds, these are 0–15 Mbps, 15–35 Mbps, 35–400 Mbps, and 400+ Mbps. For download speeds, these are 0–100 Mbps, 100–200 Mbps, 200–400 Mbps, 400–600 Mbps, and 600+ Mbps. Although we lose fine-grain analysis that may apply to specific subscriptions, we are able to perform meaningful analysis with the remaining data and avoid conflicts in categorization between testing parameters.

Once again, we find that our results and their consistencies vary across providers. We omit results from NDT7 tests because IPv4 is always labeled as faster, regardless of the speed tier. We present the dominant labels for the Ookla tests:

Charter (ASN10796): Download speeds have low normalized speeds in tier ranges 200–400 Mbps in both IP versions. Both upload and download speeds are faster in IPv4 or equivalent.
Comcast (ASN7922): Speeds are faster in IPv6 in download ranges 400–600 Mbps, and upload ranges 15–35 Mbps. Otherwise, IPv4 is faster.
AT&T (ASN7018): Labels are dependent on the Ookla speed test server used.

In households in Comcast (ASN 7922), although our labeling only requires 60% of tests to be faster for this label, we find that almost 80% of tests from each Ookla server are faster in IPv6 in the specified upload and download ranges. We hypothesize that the shifts between different speed tiers are due to differences in infrastructure for different service ranges. For households in AT&T (ASN 7018), we see that when the AT&T Ookla server is used, download speeds tend to be faster in IPv6 for higher speed ranges (200–400 Mbps or 600+ Mbps). But when the Charter or Comcast Ookla server is used, either IPv4 is faster or both IP versions are equivalent. We hypothesize that



(a) Comcast households. We observe that IPv6 speeds are more likely to be faster between 400-600 Mbps, supporting the trends observed using our taxonomy.

(b) AT&T households. We observe that when using the AT&T Ookla server, speed tests are more likely to be slower in IPv4.

Fig. 11. CDFs of Ookla download speeds from households in different providers.

this is because of a path bottleneck between the client and the speed test server that may differ between broad subscription ranges. We once again point to our description of AT&T household results in Section 4.2 and emphasize that repeated experimentation with different AT&T speed test servers indicates this is not a result of our measurements themselves.

Absolute Differences in Speed. Following our analysis of speed ranges, we look at paired tests with large *absolute* differences in speed. i.e., from the difference in their speeds in Mbps. The dominant Ookla test labels are:

Charter (ASN10796): IPv4 is faster at speed differences under 60 Mbps. Low normalized speeds occur in both lower (<20Mbps) and higher differences (>60Mbps).
Comcast (ASN7922): Households between 400-600 Mbps differing by >80 Mbps between IP versions are faster in IPv6.
AT&T (ASN7018): Paired tests with differences between 80–120 Mbps are faster in IPv6; others are faster in IPv4.

We again observe patterns per broadband provider. We also highlight that significant absolute differences in speeds are likely to affect user experience, depending on whether the online application is using IPv4 or IPv6.

Examples of Systematic Differences in Performance. With the results of our taxonomy in mind, we provide examples of our results in absolute terms (Mbps) in Figure 11, offering a direct view of the implications of our results on performance. An aggregate view of test results in each IP version does not illustrate the relationship between *paired* tests (i.e., the relationship between IPv4 and IPv6) that we can understand using relative metrics (i.e., relative speed) and our taxonomy labels. However, we use this to confirm that patterns emerge that we can understand more in-depth through taxonomizing the paired speed tests.

Through our taxonomy, we identified that Comcast households are more likely to be significantly faster in IPv6 (80+ Mbps) at speeds between 400-600 Mbps. In Figure 11a, we observe this effect in Ookla download speed tests. We observe significant gaps between IPv4 and IPv6 results in tests within this speed range.

We once again visualize a difference identified by our taxonomy labels in AT&T households, where we found that Ookla test results differ by the speed test server used. In Figure 11b, we visualize that AT&T households are more likely to diverge in speed results at higher speed tiers when using an AT&T speed test server. Here, IPv6 tests are more likely to be faster than IPv4, especially in higher speed tiers. As described in Section 4.2, we emphasize that we repeated this experiment at a much lower cadence (i.e., with 1 paired speed test every 3 hours) with different AT&T devices and a different AT&T speed test server. We observed the same relationship between IPv4 and IPv6 speed tests, and therefore do not point to individual speed test servers being overwhelmed during our measurements.

Takeaway: When considerable differences (>5%) in throughputs occur, the IP version that is faster corresponds to the provider and speed tier range. This implies that different service ranges elicit different infrastructures that are prone to faster speeds in one IP version.

5.6 Why Do Speeds Differ?

Here, we discuss some reasons why speeds in one IP version may be faster than in the other, reflecting on our earlier results.

When IPv4 is faster. IPv4 speeds may be faster because high-speed residential network infrastructure may be provisioned for IPv4 only. This optimization may exist because most toplist domains remain IPv4-only (nearly 70% according to previous work [23]). Therefore, ISPs may expect most residential traffic to be over IPv4. Furthermore, there may be more router provisioning in IPv4, and some optimizations are exclusively available in IPv4. Two examples of this are Fast Track [35], where packets are forwarded faster by bypassing features such as firewall rules, and Hardware offloading [4], where router functions are executed on hardware rather than software.

When IPv6 is faster. IPv6 speeds may be faster due to increased congestion in IPv4, but not in IPv6. Due to most toplist domains only being available over IPv4 [23], more traffic from other devices in the network may use IPv4. Furthermore, prior work identified that IPv4 paths tend to have more middleboxes than those in IPv6 [21], potentially slowing down connections.

In Section 5.3, we point to these as possible reasons for observing that the majority of households have faster IPv4 speeds and a minority have faster IPv6 speeds. In Section 5.5 **ISPs**, in Comcast households, we observed faster upload speeds in IPv4 and faster download speeds in IPv6. Upload speeds may be faster in IPv4 because there is less congestion (i.e., most client traffic is download, not upload). On the other hand, IPv6 may be faster in download speeds because there is less contention. Then, in Section 5.5 **Speed Tier Ranges** and **Examples of Systematic Differences in Performance**, we observed high differences (>80 Mbps) in Comcast households in tiers 400-600 Mbps. We emphasize that differences in speed tiers can indicate different infrastructure that experiences different bottlenecks. The fact that this does not occur in low tiers may indicate newer infrastructure that is better provisioned for IPv6. Comcast is well known to deploy native residential IPv4 [28]; it is possible that their lower tiers are optimized for IPv4 only.

Speed Test Measurement Infrastructure. We acknowledge that some differences may be due to speed test measurement configuration, specifically, the speed test server and the speed test software. While this does not identify causes of performance differences in households, this is an important consideration for future speed test measurements. In Section 5.4, we once again observed that an Ookla speed test server in AT&T reports slower IPv4 tests. In both Section 5.4 and Section 5.5 **ISPs**, we observed faster IPv4 results in NDT7 tests. Because NDT7 only uses a single TCP connection and residential routers may not offload or optimize these flows, this single IPv6 connection may fall back to CPU and software paths that limit performance.

We emphasize that observed differences in speeds are likely not due to a particular factor; they are due to multiple confounding factors. Moreover, our data is observational, and we cannot identify the true causes for differences. Rather, we show that systematic differences occur, highlighting directions for future work.

In addition to comparing throughputs directly, we also find that other factors, such as latency, intermediate path latency, and jitter, do not have a strong correlation to *differences* in speeds between IP versions. In Appendix C, we discuss our findings in more detail.

6 Concluding Discussion

Although IPv4 and IPv6 are both deployed widely in residential networks, they are not equivalent. This work compares throughputs recorded across IPv4 and IPv6 and two speed test services in residential networks. Our work presents a systematic evaluation that concludes that the IP version of the speed test has an effect on the results. Although we do not claim that our observed patterns regarding *when* speed tests differ generalize to the entirety of the Internet, we highlight similarities and differences in our measurement sample according to different conditions, including the type of software and the broadband provider. We found that, while the majority of households experience similar speeds in both IP versions in most instances, patterns emerge when we exclusively analyze tests with significant differences. These patterns emerge per broadband provider, pointing to potential bottlenecks specific to both the local network and the path between the household and the testing server. We additionally identified several patterns in speed test software.

We now make recommendations for future broadband measurement, discuss the impact of using IPv4 vs IPv6 in applications, and explore future work.

6.1 Recommendation for Broadband Speed Tests

In this work, we find several differences between IPv4 and IPv6 recorded speeds. Based on our findings, we make several recommendations for future work that measures Internet quality. We hope subsequent work can expand our measurements to more vantage points, including households with more router vendors and different broadband providers. With more conditions represented in our dataset, we hypothesize that we could identify other bottlenecks unique to each IP version.

Based on our observations, our foremost recommendation is to consider the IP version of the test in analysis by testing in both IP versions or separating results by IP version. When possible, it is important to understand the IP versions available to the client one measures from, and control for IP version in the speed test software. Failing to measure in one IP version may paint an incomplete picture of the client's network or other factors affecting performance.

Aligning with prior work [30], we encourage future work to use different types of speed test software and be conscious of the unique differences that may emerge in each based on the IP version of the connection. For example, if a lower speed is observed in tests, we encourage researchers to investigate whether this occurs in only one IP version and, if possible, repeat the test in the other IP version. As other works have shown that factors such as Wifi [53] and subscription tier [42] can affect measurement results, we hope that future work will add IP version to the factors that they consider in analysis.

6.1.1 Speed Test Software. In Section 5.4, we found that NDT7 speed test results are more likely to be faster in IPv4 compared to Ookla speed tests. The cause of this result could be due to several differences between the two types of speed tests, including both the testing software itself and the infrastructure. Aligning with prior work [30], we emphasize that the results from each software are not *right* or *wrong*; rather, they model different types of connections. As they both help us learn different things about the network, we encourage future work to use different types of speed test software and be conscious of the unique differences that may emerge in each based on the connection's IP version.

6.1.2 Speed Test Servers. In Section 4.2, we identified an Ookla speed test server that tends to report lower IPv4 download speeds compared to other servers. Prior work [30] additionally identifies inconsistencies in results related to specific Ookla servers. We similarly identified in Table 1 anomalies related to the consistency of tests using the AT&T Ookla server. We contacted Macmillan et al. and confirmed that they did not control for IP version; therefore, they have a mix of IPv4 and IPv6 results in their analysis. We hypothesize that their finding that some Ookla servers under-report speed may only have existed in one IP version. Thus, we recommend that future work consider results from multiple Ookla test servers.

6.2 IP Version Impact on User Experiences

Many dual-stack applications use Happy Eyeballs to choose the IP version. This algorithm prefers IPv6 but is otherwise based on the connection time in each protocol [52, 64]. Given our lack of correlation between latency and difference in throughput (Appendix C), we hypothesize that the fastest speed may not always be chosen. Moreover, we found that IPv4 speeds tend to be faster in Section 5.1. As Happy Eyeballs prioritizes IPv6, it is likely that a slower speed is used. We highlight the negative implications on user quality of experience if IPv4 is faster, but applications prefer IPv6.

Although it is not documented whether a Happy Eyeballs algorithm is used in Ookla and NDT7, we highlight that the IP version of the test may be different than the IP version that an application chooses. That is, a user debugging their connectivity problems in an application using *IPvX* may use a speed test that uses *IPvY*. This scenario could provide confusing results that obfuscate Internet performance. As shown in Section 5.1, we find that IP protocols degrade in speed at different times. Therefore, we encourage online application developers to consider IP version choice with respect to throughput, as opposed to latency in Happy Eyeballs. If throughput decreases, switching to another IP version may improve user experiences.

While IPv6 deployment is integral to the future of the Internet, we emphasize that adequate IPv6-specific performance testing must be conducted to maintain user experience.

6.3 Future Work

We hope subsequent work can expand our measurements to more vantage points, including households with more router vendors and different broadband providers. Furthermore, we hope that future work can incorporate other broadband connection characteristics, such as evaluating the impact of WiFi versus Ethernet, how network activity affects speeds in individual IP versions, and inferring the subscription tier of measured households [42].

7 Acknowledgements

The authors are grateful for the assistance and support of Abhishek Bhaskar. This work was supported by the National Science Foundation (NSF) Graduate Research Fellowship (GRFP) under Grant No. DGE-2039655, and Computer and Network Systems (CNS) awards 2223610, 2224687, 2319315, and 2319603. Results presented in this paper were obtained using the FLOTO instrument supported by the National Science Foundation award 2213821. The authors would like to thank the Internet Innovation Initiative and FLOTO teams for their support throughout this project, especially Marc Richardson, Mark Powers, and Jesse London. We are thankful for the anonymous reviewers and our shepherd, Gareth Tyson, for their helpful and constructive comments.

References

[1] ARRIS. Bgw210-700 broadband gateway, 2016. https://www.dslreports.com/r0/download/2346873~95c4c01f5458385e310472e8089ac04c/BGW210Manual.pdf.

[2] Bajpai, V., and Schönwälder, J. Measuring the effects of happy eyeballs. In *Proceedings of the 2016 Applied Networking Research Workshop* (New York, NY, USA, 2016), ANRW '16, Association for Computing Machinery, p. 38–44.

[3] Beltiukov, R., Chandrasekaran, S., Gupta, A., and Willinger, W. Pinot: Programmable infrastructure for networking. In *Proceedings of the 2023 Applied Networking Research Workshop* (New York, NY, USA, 2023), ANRW '23, Association for Computing Machinery, p. 51–53.

[4] Center, U. H. Edgerouter - hardware offloading. https://help.uisp.com/hc/en-us/articles/22591077433879-EdgeRouter-Hardware-Offloading.

[5] Cisco. Performance-comparison testing of ipv4 and ipv6 throughput and latency on key cisco router platforms, 2007. https://community.cisco.com/legacyfs/online/legacy/4/9/1/15376194-IPv6perf_wp1f.pdf.

[6] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

[7] Debbarma, S., and Das, A. Empirical measuring ipv4/ipv6 network performance on microsoft windows operating systems. In *2013 Third International Conference on Advances in Computing and Communications* (2013), pp. 393–395.

[8] Dhamdhere, A., Luckie, M., Huffaker, B., claffy, k., Elmokashfi, A., and Aben, E. Measuring the deployment of ipv6: topology, routing and performance. In *Proceedings of the 2012 Internet Measurement Conference* (New York, NY, USA, 2012), IMC '12, Association for Computing Machinery, p. 537–550.

[9] Dischinger, M., Haeberlen, A., Gummadi, K. P., and Saroiu, S. Characterizing residential broadband networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement* (New York, NY, USA, 2007), IMC '07, Association for Computing Machinery, p. 43–56.

[10] Feamster, N., and Livingood, J. Measuring internet speed: current challenges and future recommendations. *Commun. ACM 63*, 12 (Nov. 2020), 72–80.

[11] Federal Communications Commission (FCC). Mapping broadband health in america. Tech. rep., Federal Communications Commission, 2023. https://www.fcc.gov/health/maps.

[12] Federal Communications Commission (FCC). Measuring broadband america (mba), 2023.

[13] Federal Communications Commission (FCC). Measuring broadband raw data releases, 2023.

[14] Federal Communications Commission (FCC). Measuring fixed broadband - thirteenth report. Tech. rep., Federal Communications Commission, 2024.

[15] Fries, J., Doan, T. V., Gupta, R., and Bajpai, V. An eight years perspective on the internet broadband infrastructure in the usa. In *2022 IFIP Networking Conference (IFIP Networking)* (2022), pp. 1–9.

[16] Goel, U., Steiner, M., Wittie, M. P., Flack, M., and Ludin, S. A case for faster mobile web in cellular ipv6 networks. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2016), MobiCom '16, Association for Computing Machinery, p. 176–188.

[17] Google. Ipv6 statistics, 2024. https://www.google.com/intl/en/ipv6/statistics.html.

[18] Grey, A.-M. The case for connectivity, the new human right. *UN Chronicle* (2020). Originally published for Human Rights Day 2020.

[19] Grover, S., Park, M. S., Sundaresan, S., Burnett, S., Kim, H., Ravi, B., and Feamster, N. Peeking behind the nat: an empirical study of home networks. In *Proceedings of the 2013 Conference on Internet Measurement Conference* (New York, NY, USA, 2013), IMC '13, Association for Computing Machinery, p. 377–390.

[20] Han, J., Lyu, M., and Sivaraman, V. On the validity of internet speed test tools and broadband measurement programs. In *Proceedings of the 17th Asian Internet Engineering Conference* (New York, NY, USA, 2022), AINTEC '22, Association for Computing Machinery, p. 1–9.

[21] Hilal, F., and Gasser, O. Yarrpbox: Detecting middleboxes at internet-scale. *Proc. ACM Netw. 1*, CoNEXT1 (July 2023).

[22] Howard, L. Why is ipv6 faster?, 2019. https://blog.apnic.net/2019/07/24/why-is-ipv6-faster/.

[23] Hsu, A., Li, F., Pearce, P., and Gasser, O. A first look at nat64 deployment in-the-wild. In *International Conference on Passive and Active Network Measurement* (2024), Springer, pp. 112–129.

[24] Huston, G. Examining ipv6 performance, 2015. https://labs.ripe.net/author/gih/examining-ipv6-performance/.

[25] Keahey, K., Feamster, N., Martins, G., Powers, M., Richardson, M., Schrubbe, A., and Sherman, M. Discovery testbed: An observational instrument for broadband research. In *2023 IEEE 19th International Conference on e-Science (e-Science)*. IEEE, October 2023.

[26] Li, F., Wang, X., Pan, T., and Yang, J. A Case Study of IPv6 Network Performance: Packet Delay, Loss, and Reordering. *Mathematical Problems in Engineering 2017* (October 2017), 1–10.

[27] Li, K.-H., and Wong, K.-Y. Empirical analysis of ipv4 and ipv6 networks through dual-stack sites. *Information 12*, 6 (2021).

[28] Livingood, J. Deployment of ipv6 begins. https://corporate.comcast.com/comcast-voices/ipv6-deployment.

[29] MacMillan, K., Mangla, T., Richardson, M., and Feamster, N. Best practices for collecting speed test data. In *The Research Conference on Communications, Information, and Internet Policy (TPRC)* (2022).

[30] MacMillan, K., Mangla, T., Saxon, J., Marwell, N. P., and Feamster, N. A comparative analysis of ookla speedtest and measurement labs network diagnostic test (ndt7). *Proc. ACM Meas. Anal. Comput. Syst. 7*, 1 (Mar. 2023).

[31] Major, D., Teixeira, R., and Mayer, J. No wan's land: Mapping u.s. broadband coverage with millions of address

queries to isps. In *Proceedings of the ACM Internet Measurement Conference* (New York, NY, USA, 2020), IMC '20, Association for Computing Machinery, p. 393–419.

[32] Manda, H., Srinivasavaradhan, V., Koduru, L., Zhang, K., Zhou, X., Paul, U., Belding, E., Gupta, A., and Narechania, T. N. The efficacy of the connect america fund in addressing us internet access inequities. In *Proceedings of the ACM SIGCOMM 2024 Conference* (New York, NY, USA, 2024), ACM SIGCOMM '24, Association for Computing Machinery, p. 484–505.

[33] Measurement Lab. Measurement data, 2025. https://www.measurementlab.net/data/.

[34] Measurement Lab. Speedtest by measurement lab, 2025. https://speed.measurementlab.net.

[35] MikroTik. Manual:ip/fasttrack, 2018. https://wiki.mikrotik.com/Manual:IP/Fasttrack.

[36] Narayan, S., Shang, P., and Fan, N. Network performance evaluation of internet protocols ipv4 and ipv6 on operating systems. In *2009 IFIP International Conference on Wireless and Optical Communications Networks* (2009), pp. 1–5.

[37] NetGear Community Forum. Why ipv6 can be slower than ipv4, 2022. https://community.netgear.com/t5/Nighthawk-Wi-Fi-5-AC-Routers/Why-IPv6-can-be-slower-than-IPv4/td-p/2231369.

[38] Nikkhah, M., Guérin, R., Lee, Y., and Woundy, R. Assessing ipv6 through web access a measurement study and its findings. In *Proceedings of the Seventh COnference on Emerging Networking EXperiments and Technologies* (New York, NY, USA, 2011), CoNEXT '11, Association for Computing Machinery.

[39] Ookla. Speedtest by ookla, 2024. https://www.speedtest.net.

[40] Ookla. Speedtest by ookla global fixed and mobile network performance maps, 2024. https://registry.opendata.aws/speedtest-global-performance/.

[41] Paul, U., Gunasekaran, V., Liu, J., Narechania, T. N., Gupta, A., and Belding, E. Decoding the divide: Analyzing disparities in broadband plans offered by major us isps. In *Proceedings of the ACM SIGCOMM 2023 Conference* (New York, NY, USA, 2023), ACM SIGCOMM '23, Association for Computing Machinery, p. 578–591.

[42] Paul, U., Liu, J., Gu, M., Gupta, A., and Belding, E. The importance of contextualization of crowdsourced active speed test measurements. In *Proceedings of the 22nd ACM Internet Measurement Conference* (New York, NY, USA, 2022), IMC '22, Association for Computing Machinery, p. 274–289.

[43] Peck, R., Olsen, C., and Devore, J. *Chapter 11: Comparing Two Populations or Treatments*, fourth ed. 2000.

[44] Pujol, E., Richter, P., and Feldmann, A. Understanding the share of ipv6 traffic in a dual-stack isp. In *Passive and Active Measurement: 18th International Conference, PAM 2017, Sydney, NSW, Australia, March 30-31, 2017, Proceedings 18* (2017), Springer, pp. 3–16.

[45] Reddit. Ipv6 is 2x as fast as ipv4, 2021. https://www.reddit.com/r/tmobileisp/comments/sj41oo/ipv6_is_2x_as_fast_as_ipv4/.

[46] Reddit. I'm out of the loop with the ipv6 issues. been having slow speeds for a few months, decided to upgrade to 1gbs speed, but still seeing mediocre upload rates. what's the deal?, 2022. https://www.reddit.com/r/Fios/comments/13v7hz2/im_out_of_the_loop_with_the_ipv6_issues_been/.

[47] Richter, P., Allman, M., Bush, R., and Paxson, V. A primer on ipv4 scarcity. *SIGCOMM Comput. Commun. Rev. 45*, 2 (Apr. 2015), 21–31.

[48] Richter, P., Wohlfart, F., Vallina-Rodriguez, N., Allman, M., Bush, R., Feldmann, A., Kreibich, C., Weaver, N., and Paxson, V. A multi-perspective analysis of carrier-grade nat deployment. In *Proceedings of the 2016 Internet Measurement Conference* (New York, NY, USA, 2016), IMC '16, Association for Computing Machinery, p. 215–229.

[49] RIPE NCC. RIPE Atlas, 2025. https://atlas.ripe.net/.

[50] Sargent, M., and Allman, M. Performance within a fiber-to-the-home network. *SIGCOMM Comput. Commun. Rev. 44*, 3 (July 2014), 22–30.

[51] Sattler, P., Kirstein, M., Wüstrich, L., Zirngibl, J., and Carle, G. Lazy eye inspection: Capturing the state of happy eyeballs implementations. *arXiv preprint arXiv:2412.00263* (2024).

[52] Schinazi, D., and Pauly, T. Happy Eyeballs Version 2: Better Connectivity Using Concurrency. RFC 8305, Dec. 2017.

[53] Sharma, R., Feamster, N., and Richardson, M. A longitudinal study of the prevalence of wifi bottlenecks in home access networks. In *Proceedings of the 2024 ACM on Internet Measurement Conference* (New York, NY, USA, 2024), IMC '24, Association for Computing Machinery, p. 44–50.

[54] Sharma, R., Mangla, T., Saxon, J., Richardson, M., Feamster, N., and Marwell, N. P. Benchmarks or equity? a new approach to measuring internet performance. *A New Approach to Measuring Internet Performance (August 3, 2022)* (2022).

[55] Sharma, T., Mangla, T., Gupta, A., Jiang, J., and Feamster, N. Estimating webrtc video qoe metrics without using application headers. In *Proceedings of the 2023 ACM on Internet Measurement Conference* (New York, NY, USA, 2023), IMC '23, Association for Computing Machinery, p. 485–500.

[56] Sharma, T., Marques, J., Marwell, N., and Feamster, N. A first look at the spatial and temporal variability of internet performance data in hyperlocal geographies. In *The Research Conference on Communications, Information, and Internet Policy (TPRC)* (2023).

[57] Smith, M. The trouble with nat: Part 3, 2016. https://blog.apnic.net/2016/10/27/trouble-nat-part-3/.

[58] Sundaresan, S., Burnett, S., Feamster, N., and de Donato, W. BISmark: A testbed for deploying measurements and applications in broadband access networks. In *2014 USENIX Annual Technical Conference (USENIX ATC 14)* (Philadelphia, PA, June 2014), USENIX Association, pp. 383–394.

[59] Sundaresan, S., de Donato, W., Feamster, N., Teixeira, R., Crawford, S., and Pescapè, A. Broadband internet performance: a view from the gateway. *SIGCOMM Comput. Commun. Rev. 41*, 4 (Aug. 2011), 134–145.

[60] Sundaresan, S., Feamster, N., and Teixeira, R. Measuring the performance of user traffic in home wireless networks. In *Passive and Active Measurement* (Cham, 2015), J. Mirkovic and Y. Liu, Eds., Springer International Publishing, pp. 305–317.

[61] The ZMap Project. Zgrab 2.0, 2024. https://github.com/zmap/zgrab2.

[62] UChicago FLOTO. Ipv4 and ipv6 speed tests, 2025. https://github.com/UChicago-FLOTO/ipv4-ipv6-speedtests.

[63] Wang, Y., Ye, S., and Li, X. Understanding current ipv6 performance: a measurement study. In *10th IEEE Symposium on Computers and Communications (ISCC'05)* (2005), pp. 71–76.

[64] Wing, D., and Yourtchenko, A. Happy Eyeballs: Success with Dual-Stack Hosts. RFC 6555, Apr. 2012.

[65] Zhou, X., Jacobsson, M., Uijterwaal, H., and Van Mieghem, P. Ipv6 delay and loss performance evolution. *International Journal of Communication Systems 21*, 6 (2008), 643–663.

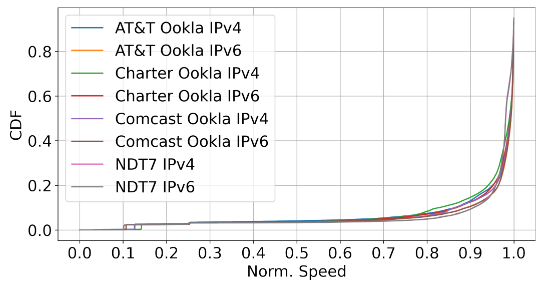


Fig. 12. CDF of the normalized speeds from all speed tests.

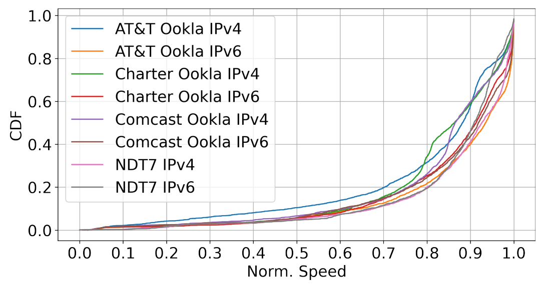


Fig. 13. CDF of the normalized speeds from speed tests with a relative difference of over 5%.

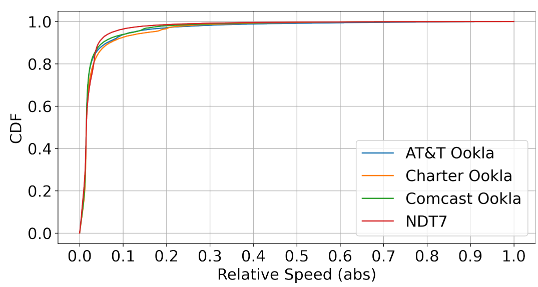


Fig. 14. CDF of relative speeds from paired speed tests.

A Thresholds in Taxonomy and Filtering

Taxonomy. We choose to use a normalized speed value of 0.8 or below as "low" in our taxonomy. To illustrate the reason for this choice, we show the distribution of all normalized speeds in our dataset in Figure 12. We observe that 0.8 is slightly below the modal shift of this plot, indicating it is slightly below the normalized value at which speeds begin to decrease in performance.

We additionally observe that when we plot all our *filtered* (relative speed difference > 5%) in Figure 13 there is a better-defined modal shift around 0.8. We use this as motivation to choose 0.8 as our threshold for a "low" normalized speed in our taxonomy described in Section 5.2.

Filtering for Relative Speed Differences. We use a relative speed difference of 5% in Section 5.4. Illustrated in Figure 14, we choose this as our filtering threshold because we see a modal shift around 0.05.

	AT&T Ookla	Charter Ookla	Comcast Ookla	NDT7
Down	11.9%	13.3%	10.5%	8.5%
Up	12.8%	12.7%	13.6 %	18.3%

Table 4. The percentage of speed tests that have a relative speed difference of 5% or more per speed test type.

B Speed Test Data After Filtering

As described in Section 5.4, after filtering for considerable differences in relative speed (>5% difference between IP versions), we are left with a subset of our data. Table 4 shows the number of tests that remain after this filtering in all speed test parameters.

C Independence of Speed Variations to Other Network Characteristics

Contrary to expectation, we find that many network attributes measured alongside the speed tests have no observable pattern to *differences* in IPv4 and IPv6 speeds. We conduct several other measurements alongside our speed tests and use additional data that the speed test software collects. In both IPv4 and IPv6, we analyze:

- The city of the household.
- Latency to the speed test server.
- Jitter (for Ookla, NDT7 does not provide this data).
- The router vendor.
- The day of the week.
- The time of day.
- Traceroutes to the speed test server from the device executed after the speed test.
- Latency to the household's router to estimate whether the router itself is under high load.

For all factors except for city, router vendor, day of the week, and time of day, we normalize the values per household by dividing by the 95th percentile value for that metric for that household (i.e., what we could expect this value to be). For our traceroutes, we preprocess them and use the RTT to various points including the local router, leaving the subnet of the client, and leaving the client's ASN. We also look at the total hops in the traceroute.

We use a correlation coefficient to understand whether there is a relationship between these factors and speed differences. The correlation coefficient ranges from -1 to +1, where a high absolute value indicates a linear relationship and a value close to 0 indicates little to no relationship. We calculate the correlation coefficient between our other measurements and: 1. Relative IPv4/IPv6 speed, 2. Normalized speed in IPv4 and IPv6. By using relative speed and normalized speeds, we focus on factors that predict *changes* in speeds, relative to the expectation for that household.

In almost all cases (in all Ookla and NDT7 tests across both upload and download), we find that the absolute value of the correlation coefficient is under 0.3. This is true when using the entire dataset as well as our filtered dataset (>5% difference in relative speed). This indicates that the relationship between these factors and speed and relative speed is weak, even when only looking at cases where the relative speed difference is high [43].

We find two exceptions with the router vendor and the city of the household correlated to the download IPv6 normalized speeds. The household's router vendor with the normalized IPv6 speed in filtered Charter and Comcast Ookla speed tests has correlation coefficients of 0.38 and 0.34, respectively, indicating that there is still a relatively weak relationship. Moreover, we find that the router vendor is likely to be similar among households in the same ASN. For example, every household in Comcast (ASN 7922) uses an Xfinity router. Therefore, this aligns with our existing conclusion that ASN correlates strongly with speed patterns. Similarly, in all Ookla download IPv6 normalized speeds, the correlation coefficient to the city is just over 0.3 (-0.34, -0.33, -0.37 for the AT&T server, Charter server, and Comcast servers, respectively). We emphasize that this is still a weak relationship, and the city and ASN are highly related because all of the ASN 7922 and ASN 10796 households are in different cities.

We hypothesize that future work that expands measurements to a more diverse set of households (particularly, those with more router vendors) may find a relationship between the router hardware and speed test results.

Takeaway: Several other network characteristics bear little-to-no correlation with differences in speed by IP version. We, however, note that our household demographics limit this analysis and a more diverse study is warranted.

C.1 Analysis Method Considerations

We note that we explore other analysis methods, including regression testing and clustering, to identify other patterns in speed differences, but find little success. Here, we explore such methods. Although we did not find relationships, we note that we did not exhaustively explore methodologies that model high-dimensional relationships. We encourage future work to explore these as possible avenues for more sophisticated relationships impacting differences in speed across IP versions.

We additionally experiment with other techniques to attribute differences in speed to specific network characteristics, including regression testing and clustering with Principal Component Analysis (PCA). However, we find little success when attempting these methods on the unfiltered and filtered (relative speed differences >5%) datasets. We attribute this to the *steady-state* behavior that occurs in a large fraction of our data, as well as the lack of diversity from different providers and router vendors. In regression testing, we had low R-squared values (<0.2) that indicate a poor goodness-of-fit of the model. We experimented with predicting the relative speed and normalized speeds in each IP version. In clustering methods, we applied PCA and then experimented with a large range of cluster numbers (1-20). However, we had low silhouette scores (<0.4) in all cases. Clusters also frequently separated speed tests by the broadband provider, which we already identify as an indicator of patterns in which IP version is faster.

Received July 2025; revised September 2025; accepted October 2025