The Importance of Contextualization of Crowdsourced Active Speed Test Measurements

Udit Paul
Department of Computer Science
UC Santa Barbara
u_paul@ucsb.edu

Jiamo Liu
Department of Computer Science
UC Santa Barbara
jiamoliu@ucsb.edu

Mengyang Gu
Department of Statistics and Applied Probability
UC Santa Barbara
mengyang@ucsb.edu

Arpit Gupta
Department of Computer Science
UC Santa Barbara
arpitgupta@ucsb.edu

Elizabeth Belding
Department of Computer Science
UC Santa Barbara
ebelding@ucsb.edu

ABSTRACT

Crowdsourced speed test measurements, such as those by Ookla® and Measurement Lab (M-Lab), offer a critical view of network access and performance from the user's perspective. However, we argue that taking these measurements at surface value is problematic. It is essential to contextualize these measurements to understand better what the attained upload and download speeds truly measure. To this end, we develop a novel Broadband Subscription Tier (BST) methodology that associates a speed test data point with a residential broadband subscription plan. Our evaluation of this methodology with the FCC's MBA dataset shows over 96% accuracy. We augment approximately 1.5M Ookla and M-Lab speed test measurements from four major U.S. cities with the BST methodology. We show that many low-speed data points are attributable to lower-tier subscriptions and not necessarily poor access. Then, for a subset of the measurement sample (80k data points), we quantify the impact of access link type (WiFi or wired), WiFi spectrum band and RSSI (if applicable), and device memory on speed test performance. Interestingly, we observe that measurement time of day only marginally affects the reported speeds. Finally, we show that the median throughput reported by Ookla speed tests can be up to two times greater than M-Lab measurements for the same subscription tier, city, and ISP due to M-Lab's employment of different measurement methodologies. Based on our results, we put forward a set of recommendations for both speed test vendors and the FCC to contextualize speed test data points and correctly interpret measured performance.

ACM Reference Format:
Udit Paul, Jiamo Liu, Mengyang Gu, Arpit Gupta, and Elizabeth Belding. 2022. The Importance of Contextualization of Crowdsourced Active Speed Test Measurements. In *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22), October 25–27, 2022, Nice, France.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3517745.3561441

IMC '22, October 25–27, 2022, Nice, France

ACM ISBN 978-1-4503-9259-4/22/10.
https://doi.org/10.1145/3517745.3561441

1 INTRODUCTION

The challenge of mapping fixed broadband Internet access was brought to the forefront during the stay-at-home orders of the Covid-19 pandemic. Suddenly, individuals without high-quality Internet access could not participate in the remote schooling, work, and telehealth that these orders required [1, 2, 4]. Further, federal money for Internet infrastructure improvement was made available through the Bipartisan Infrastructure Investment and Jobs Act [14]; however, a key challenge remained: knowing where high-quality Internet access was lacking [5, 19]. While the Federal Communications Commission (FCC) has long compiled annual Broadband Reports that map provider-reported access at the census block level, these reports are known to overstate access availability and speed, particularly in rural and under-served urban areas [7, 25, 42].

Crowdsourced active network measurements have emerged as a powerful tool to map fixed broadband access more accurately. These "speed tests" provide a critical snapshot of the network state from the vantage point of the end users. Because they are active measurements, they provide data on actual performance instead of the theoretical maximum performance reported by the providers. Popular network speed test platforms, such as Ookla's speedtest.net [22], Measurement Lab's speed.measurementlab.net [21], FAST [10] and Xfinity's speed test [30], are utilized by Internet users worldwide to conduct these measurements. For instance, Ookla claims over 40 billion user-initiated tests since its inception [24]. Because of the inherent benefits, numerous governmental initiatives (e.g. [6, 8, 13, 15, 29, 34]) have come to rely on crowdsourced speed test data to map broadband access. With this data, local governments, community organizations, and others can attempt to discern where to make the economic investment in infrastructure to address digital inequality. *Perhaps most critically, the FCC itself has recently specified a challenge process [11], whereby individual users and communities can gather active measurement data to challenge provider-reported coverage claims.*

However, despite the broad use of crowdsourced active network measurements and the call for their usage by the FCC, the data generated through these speed tests suffer from several key limitations, which must be addressed before drawing meaningful conclusions about fixed Internet performance. More concretely, *we argue that speed test measurements must be contextualized to accurately interpret the measured performance.* The challenge here is understanding *what a speed test measures* and *how it compares to expected speed values*. For example, many fixed broadband plans offer rates as high as 1 Gbps download and 35 Mbps upload. If a speed test measures performance significantly less than these values, is it because the access network is under-performing, the user has purchased a lower-tier plan, or the user's home WiFi network is misconfigured or experiencing interference? It is critical to determine the source of the under-performance. If the under-performance is attributable to issues in the access network, then the problem could be reported to the Internet Service Provider (ISP) to challenge coverage claims in an area. In contrast, if the under-performance is attributable to local factors, such as channel interference or poor signal quality, the user can address it directly. If the user simply purchased a lower-tier plan, then perhaps the speed test is measuring the paid-for speed. Finally, the methodology of the test itself can impact performance results, adding anothe layer of complexity [37, 39].

In this paper, we utilize more than 1.5M total measurements from Ookla and M-Lab speed tests to demonstrate the critical need for contextualization of these measurements. We start with an analysis of aggregate performance, as represented by this data, across four major metropolitan cities in the U.S. To demonstrate the importance of subscription plan context, we propose a novel approach called the Broadband Subscription Tier (BST) methodology that determines, with over 96% accuracy, the subscription plan associated with a group of speed test measurements. We evaluate the accuracy of this methodology on over 60k Measuring Broadband America (MBA) data points, for which we have subscription ground truth. After applying the methodology to our M-Lab and Ookla datasets, we show that the majority of the speed tests in a city originate from the lower subscription tiers. This implicit bias in the data skews the overall results for metrics such as download speed to lower throughputs.

Second, we incorporate the subscription tier context to Ookla measurements to quantify the impact of factors such as access type (WiFi vs. Ethernet), WiFi spectrum band, RSSI and device memory. We find that side effects of these local factors can lead to performance that only achieves half the data rate of the subscribed plan. We also evaluate the impact of the time of the test on the measured performance and, interestingly, discover minimal impact. Finally, we evaluate the performance of M-Lab versus Ookla speed test results for each subscription tier and demonstrate that M-Lab tests consistently achieve lower download speeds than Ookla tests, at times by as much as a factor of two.

In summary, our work makes the following contributions:

- We develop a novel methodology (BST) that maps crowdsourced speed test results to the residential broadband subscription plan at the test location. We demonstrate over 96% accuracy on 60k MBA data points, for which we have ground truth.
- We apply this methodology to 1.5*M* Ookla and M-Lab speed test measurements in four U.S. cities and show that the majority of data points originate from lower subscription tiers, thereby skewing throughput results.
- We quantify the impact of access type, WiFi characteristics, device memory, and time of day on Ookla measurements.

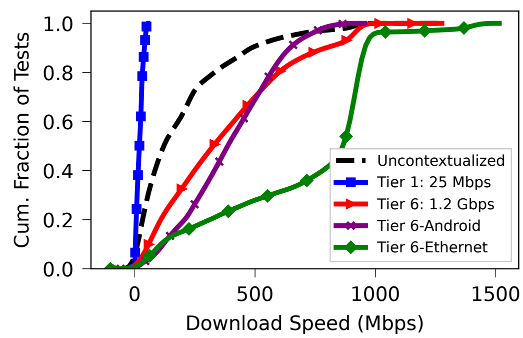


Figure 1: Comparison of raw speed test download speed distributions in a major U.S. city. The "Uncontextualized" line represents our starting point. The other lines represent the original data contextualized with subscription tier, access link speed or type, and/or device type.

- We quantify the performance difference of Ookla versus M-Lab measurements for the same subscription tiers, cities, and ISPs that stem from the differing measurement methodologies.
- Based on our results, we put forth a set of recommendations for speed test vendors and the FCC to contextualize speed test data and correctly interpret measured performance.

2 A MOTIVATING EXAMPLE

We begin by illustrating the challenge and inaccuracy of interpreting crowdsourced active measurement (e.g. "speed test") data at face value. We base our initial analysis on 745k Ookla measurements from the primary fixed broadband ISP in four major U.S. cities during 2021. The median download speed of each of these four cities is roughly 115 Mbps. In prior work, a similar analysis, emphasizing the median value of the aggregated tests, was used to study the regional Internet quality of a congressional district in New York [34]. Based on these median performance results, the report recommended regions for Internet buildout and funding allocations to improve Internet quality in the constituency.

However, as our paper will illustrate, the lack of context for these measurements prevents proper interpretation of such aggregate results. Figure 1 presents the distributions of the download speed in City-A disaggregated by subscription plan tiers, access speed or link type, and measurement device type. The "uncontextualized" line represents the original data without context applied. The figure shows that the median download speed of the lowest (slowest) subscription tier (Tier 1, with a maximum download speed of 25 Mbps) is 19.22 Mbps, almost six times as slow as the overall City-A median download speed. City-A's median download speed, on the other hand, is nearly four times less than the premium ISP subscription tier (Tier 6: 1.2 Gbps) and almost seven times less than that recorded by test takers on Tier 6 Ethernet connections (Tier 6: Ethernet). Similarly, for speed tests that do not experience local bottlenecks (tests whose performance is constrained by local WiFi factors such as WiFi band and RSSI), the median download speed of the highest subscription tier for this group of speed tests (Tier 6: Android) is

Table 1: Number of measurements for datasets utilized in this work. Note that for Ookla and M-Lab, the data points are from each city, whereas for MBA, the data points are from the state that corresponds to each city.

City/State	ISP	Ookla	M-Lab	MBA
A	1	214 k	113 k	25.9 k
B	2	205 k	376 k	14.9 k
C	3	128 k	64 k	10.9 k
D	4	198 k	166 k	8.9 k

almost four times more than the City-A median download speed. Still, the median for the group of tests not affected by local bottleneck factors is half the Tier 6 (Ethernet) median download speed rate.

In the remainder of this paper, we describe the contributions that enable us to contextualize each measurement point with broadband plan subscription tier, local network characteristics, device context, test time, and speed test vendor. In so doing, we demonstrate that the ability to contextualize speed test measurements is critical for interpreting the quality of the Internet in a region.

3 DATASETS

This section describes the three primary datasets we utilize for this work. Table 1 summarizes the number of data points of each type. We choose Ookla's Speedtest® (obtained from the Speedtest Intelligence® portal) as it is the largest Internet measurement vendor that is capable of measuring available bandwidth with high accuracy [48]. M-Lab's Speed Test, on the other hand, makes collected data publicly available. We utilize the Measuring Broadband America (MBA) dataset because it provides the subscriber's purchased broadband plan information with the speed test measurements.

We use the Ookla and M-Lab data collected from January 1 – December 31, 2021. MBA data is also from this period but lacks data from September 1 – October 31 (this data is unavailable from the MBA website).

3.1 Ookla's Speedtest

Ookla's Speedtest[1] (data provided through Ookla's Speedtest Intelligence®) possesses over 16k measurement servers worldwide [26] and allows users to assess the quality of their Internet connection using either a web-based portal or native mobile application [22]. For each Speedtest, a nearby test server is selected and *multiple TCP connections* are used to calculate the throughput of the connection. Ookla's Speedtest Intelligence dataset contains individual Speedtest measurements that include QoS metrics (up/down throughput, latency, packet loss, jitter), as well as meta-features such as ISP, device type, and access type. Ookla provides performance data aggregated over time and space to the public [23].

A Data Usage Agreement (DUA) with Ookla provides us access to over 745k individual Speedtest measurements from four major metropolitan cities in the U.S, which we use for this study. Each of these cities has a population in the range of 400,000 – 700,000. For each city, we utilize the FCC Form 477 dataset [12] to identify the dominant ISP and conduct our analysis. Specifically, we use this dataset to compute the number of census blocks served by an ISP in a city and pick the one that covers the highest number of blocks.

The Ookla dataset tags the origin of each test, specifying whether the test was initiated through a web-based portal or a native application. The web-based tests do not provide device-related information. On the other hand, the native application dataset indicates the type of device that started each measurement (Android, iOS, or desktop). 394k of the measurement points in our dataset originated from native applications. The dataset also contains critical metadata related to the wireless link for Android devices, such as frequency band, signal strength, maximum achievable theoretical downlink throughput, and available kernel memory. These metrics are essential in contextualizing the measurements, as we will show in section 6.

3.2 M-Lab's Speed Test

M-Lab's Speed Test[2] (note the different spelling and capitalization from Ookla's Speedtest) has available over 500 well-provisioned servers worldwide to conduct free performance measurement tests [21]. We utilize M-Lab's Speed Test data that reports client upload/download speed performance using the Network Diagnostic Tool (NDT). This tool establishes a *single TCP connection* to quantify uplink/downlink speeds. As a result of the single TCP connection, it often under-reports the connection capacity [37]; we quantify this under-measurement by comparing its results with tests from Ookla in section 6.3. For each test, the M-Lab data also reports the client and server IP addresses, Autonomous System Number (ASN), and round trip time (RTT). However, this dataset does not provide additional context, such as device type or features. We extracted 717k NDT measurements from the same four major U.S. cities in 2021 for the same major residential broadband ISPs as Ookla. Because NDT measurements do not associate an upload speed test with a download speed test initiated by the same client, we adopt a similar methodology to [46]. We compute a 120 second window for every download speed test and filter all upload speed tests issued from the same client and server IP address. If a single upload speed is captured during that window, we associate it with the download speed. In the event we observe more than one upload speed test started during this time frame that meets this criterion, we associate the earliest upload speed test with the download speed test. As a result, *our methodology enables us to compare Ookla and M-Lab measurements over the same period, in the same cities, for the same service provider.*

3.3 Measuring Broadband America

Measuring Broadband America (MBA) [17] is an FCC-sponsored project that uses specialized hardware test units [28] to collect Internet measurement data from 4,000 U.S. households. These units measure and report upload and download speed multiple times per day [18]. Each device in the dataset also reports its location (at the granularity of census tract). Most critically, this dataset is generated from wired devices and contains the broadband plan subscription of the user hosting the device. Wired devices provide

[1] http://speedtest.net

[2] https://speed.measurementlab.net/#/

measurement data of the access link without confounding WiFi performance, while the broadband plan data provides ground-truth for our methodology to determine broadband subscription tier. We utilize the latest subscriber information, which was collected in 2020, for the measurements [27].

3.4 Ethics

While our work analyzes speed tests from users of two prominent speed test vendors, our work is not human subjects research. The private dataset shared by Ookla under DUA is fully anonymized, and we cannot identify the individual users of the platform. For the subset of measurements from devices with GPS geolocation enabled, Ookla provides GPS coordinates truncated after three decimal points. Such geolocation is accurate to 111 metres; therefore, we cannot associate it with any user/residence. The M-Lab dataset provides only public IP addresses that one can localize using IP geolocation tools. However, IP geolocation errors can exceed 30 KM, making it difficult to isolate specific users/homes. We also obtained the street address dataset from Zillow under a DUA. We do not have methods to identify residents, selected broadband subscription tiers, or the actual speed test performance at any address.

4 DETERMINING SUBSCRIPTION TIERS

Our first step in contextualizing speed test data is to determine the home broadband subscription tier of the user from which the measurement originates. This step is critical because it provides context for the achieved download and upload speed; with information about the theoretical maximum speeds (the "plan" speeds), we can first determine whether a speed test measurement indicates the network is under-performing. Without this information, we may attribute a slow download speed to the under-performance of the access link instead of a lower ("slower") tier plan purchased by the user.

To determine the subscription tier, we must first obtain the residential broadband plans available at the location of the speed test so that we know the set of possible plans from which to select. As described in this section, we obtain this information by modifying a prior approach. Then, we apply our Broadband Subscription Tier (BST) methodology, a novel two-stage hierarchical unsupervised clustering technique that matches each <download speed, upload speed> measurement tuple to a specific subscription plan.[3] To evaluate the efficacy of BST, we utilize the MBA dataset as it provides both the speed test measurements and subscription tier information for more than 60k data points.

Challenges. There exist two significant challenges in associating crowdsourced measurements with subscription tier information. First, no dataset exists in the public domain that details all the broadband plan choices offered by ISPs to users at the granularity of street address, census block, or even census block group. Through its Form 477 [12], the FCC only provides the ISP-reported maximum download/upload speed in a census block. Unfortunately, it is impossible to associate measurements with subscription tiers without a complete picture of all the plans available from the ISPs. Second, crowdsourced measurement results are inherently noisy, as they are vulnerable to environmental factors that range from poor WiFi router positioning to device memory, as shown in section 6.1. As such, it is crucial to understand the variability between different metrics reported through speed tests prior to assigning a measurement to a subscription tier.

[3] We use the terms "subscription tier" and "subscription plan" interchangeably.

4.1 Observations

To obtain the set of ISP-offered subscription plan choices, we modify the tool proposed in [42]. In particular, we augment the tool to collect available download/upload speed plans for major residential ISPs at specific U.S. street addresses. Our tool requires clean and well-formatted street addresses to obtain this information. Hence we utilize the residential property address dataset from Zillow [31] to create an address set for each of the four cities in our study. Then, we randomly select 100K residential addresses for each city and collect the ISP-offered plans. To prevent overloading ISP infrastructure, we carefully limit the number of queries we make per ISP. Our analysis of street-address level broadband plan choices in four cities reveals two significant trends.

The first trend we observe is that the plan choices remain unchanged across different street addresses within a city. For example, ISP-A offers six plans for all street addresses in City-A. Three of these plans have different download speeds (25 Mbps, 100 Mbps, and 200 Mbps) but the same upload speed (5 Mbps). The other three plans have different, faster download speeds (400 Mbps, 800 Mbps, and 1200 Mbps) with upload speeds of 10 Mbps, 15 Mbps, and 35 Mbps, respectively. We observe similar types of tiered offered plans that do not vary based on the specific address for the other three cities and major ISPs.

Second, although an ISP offers diverse plans for download speeds, varying in both number and speed range, the set of maximum available upload speeds is much smaller. Further, the upload speeds are much slower than available download speeds. This observation is noteworthy because, as discussed in [39], many factors, such as local home network conditions (e.g., WiFi interference or local congestion) and web-browser limitations, could prevent a speed test measurement from attaining high throughput. On the other hand, given the lower maximum upload speeds, fewer factors can limit the attainment of the maximum speeds [47].

As a result, crowdsourced measurements from individual users should exhibit less variation (and more consistency) in upload speed compared to download speed. Given this intuition about upload speed, we should expect to see that the recorded upload speeds during multiple measurements for a single user are more consistent than the set of download speeds for that user. To capture this per-user performance consistency, we calculate a consistency factor by taking the ratio of the mean and 95^{th} percentile for the sets of upload and download speeds recorded over multiple tests by the same user [44]. The closer the consistency factor is to 1, the greater the consistency for the evaluated metric over the set of tests from a single user.

Concretely, we select measurements from any Ookla user who conducted at least five tests using the native application while connected to the WiFi network [44]. In total, 23k (out of 85k) users issued more than five tests. These users contribute 80k measurements, about 70% of total measurements from native applications.

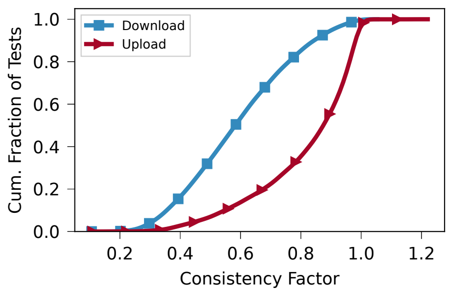


Figure 2: CDF of consistency factor for all iOS users who recorded at least five tests.

For brevity, we present the results only for City-A. We base our analysis only on native app users because a public IP address identifies users of web-based tests. Given the prevalence of NAT employed by the ISPs, determining which group of tests belongs to an individual user based on the public IP address is highly challenging.

Figure 2 shows the CDF of the consistency factor of measurements from users who registered at least five tests using Ookla's native iOS application (we present only the iOS result for clarity and confirm that we observed similar trends for data from Android and desktop native applications). As shown in the figure, download speed variations are much more significant than upload speed; upload speed is more consistent across all users. The median consistency factor for download speed is 0.58, compared to 0.87 for upload speed. The more consistent behavior of upload speed performance indicates the possibility of utilizing this metric to determine the subscription tier for each speed test. We confirm our observations of upload speed consistency for the other three cities. Note that while we report the mean value, we do observe that the consistency factor exceeds one for some users. The mean value of a (heavy-tailed) distribution can be skewed by larger items in the tail portion of the distribution.

Combining these two observations, *we hypothesize that we can utilize the measured upload speeds of the speed tests to identify the subset of possible subscription plans from which any given speed test originates.* In the next section, we describe our Broadband Subscription Tier methodology, which is our approach to matching speed test measurements to their corresponding subscription plan.

4.2 BST methodology

We propose a two-stage hierarchical unsupervised clustering methodology to match each $< download\ speed, upload\ speed >$ measurement tuple to a specific ISP subscription plan. In the first stage, our objective is to associate the recorded upload speed of a speed test to a cluster that corresponds to the correct ISP-offered upload speed. Because multiple plans might offer the same upload speed, in the second stage we use our first stage clustering to perform an inter-cluster analysis to identify the set of individual subscription tiers to which a recorded download speed can potentially match. Combining the two stages yields a probabilistic model that can map the results of speed test measurements to their respective subscription classes/tiers. Figure 3 gives an overview of our methodology.

Table 2: BST upload speed selection accuracy for the four states in the MBA dataset.

State	ISP	#Units	Accuracy
A	1	20	99.33%
B	2	17	98.19%
C	3	10	96.84%
D	4	11	99.10%

For a given speed test dataset in a city, each of our two stages begins by first confirming the presence of clusters within the upload/download speed distribution. Taking the example of the first stage, we start by employing a Kernel Density Estimation (KDE) [16] method with multivariate Gaussian kernel functions to estimate the probability densities of the upload speeds recorded during the speed tests. Combining these multiple kernel functions results in a smooth function that produces clusters containing the upload speed densities. This stage checks whether the number of upload/-download speeds offered by an ISP matches the number of clusters formed in the distribution of crowdsourced measurements.

After determining the number of clusters using the KDE method, we cluster the upload speeds by employing the Gaussian Mixture Model (GMM) [20] to determine the upload speed of the subscription tier. Once a measurement is associated with a cluster of upload speed, we enter the second stage, where we re-apply GMM to determine the corresponding download speed cluster. Note here that we possess the information about the mapping between different offered download and upload speeds through the mechanism described in section 4.1.

We choose GMM because it is one of the most popular unsupervised clustering techniques employed on a distribution consisting of several components of Gaussian densities. In GMM, each cluster follows a Gaussian distribution, and the eventual goal is to assign measurements to different parts by estimating each cluster's parameters. The parameters associated with a GMM cluster/component include the mean, covariance matrix, and weight. As such, compared to other clustering methodologies such as K-Means, GMM is a probabilistic model that considers the clusters' variance in addition to the means. In each stage, we employ GMM in conjunction with the Expectation-Maximization (EM) [9] methodology (GMM-EM) to iteratively compute the maximum likelihood that each speed test data point belongs to its respective upload/download speed cluster.

4.3 Evaluation with MBA dataset

We leverage the MBA dataset to evaluate the efficacy of our BST methodology. This dataset contains not only active measurements collected hourly but also subscription information. We apply our BST methodology to 60k measurements in this dataset spanning the four states associated with the four cities in our study. We compare the result of BST with the ground truth subscription information available in the MBA dataset by calculating the $accuracy = (\frac{\#correctly\ associated\ measurements}{\#total\ measurements})$. Table 2 presents the total number of units and the corresponding accuracy achieved

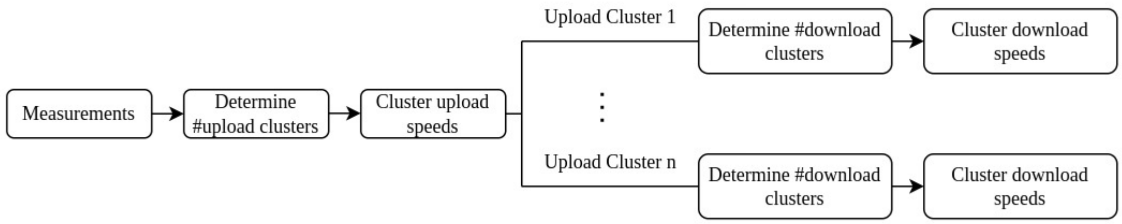


Figure 3: Broadband Subscription Tier methodology.

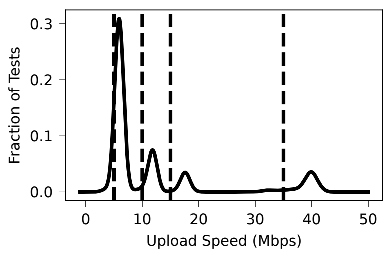


Figure 4: Upload speed density using KDE method on MBA State-A dataset. The vertical lines are the upload speed plans offered by ISP-A.

by the BST methodology for upload speeds. For all states, accuracy is above 96%; accuracy is above 99% for two states.

As a descriptive example, we provide a detailed explanation of the application of the BST methodology to the MBA dataset in State-A, where ISP-A is the dominant residential Internet service provider. Table 2 shows that 20 measurement units subscribe to ISP-A in this state. These units record a total of 25, 927 measurements during 2021. The plans recorded for the MBA subscribers in State-A are similar to the offered plans described previously for City-A. However, there are no records of the 25 Mbps download (5 Mbps upload) subscription plan in the MBA-State-A dataset. This observation is important when we match subscription plans to measurements in the following example.

Upload Speed Subscription Tiers. We begin by applying KDE on the set of upload speeds measured by the MBA nodes in State-A; Figure 4 presents the result. There are four significant clusters of upload speed densities in this dataset. The distinct peaks of upload speed densities in the regions of the offered upload speeds by ISP-A indicate the possibility of identifying the subscription plan of a given measurement.

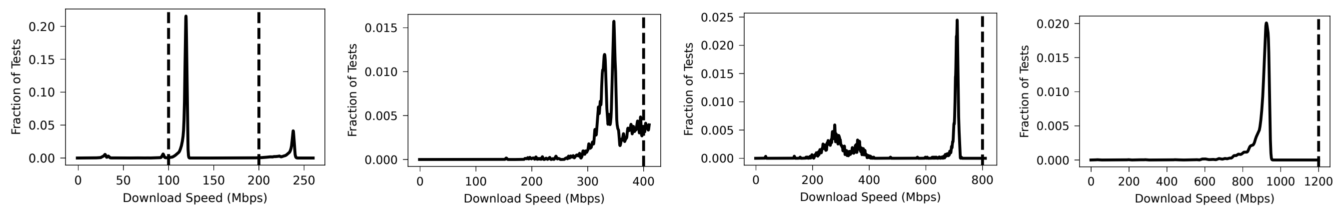
After determining the number of clusters, we aim to assign each measurement point to the appropriate subscription tier by first using the recorded upload speed. To do so, we employ the BST methodology to detect the clusters of the upload speed recorded by the MBA units. The methodology converged after 20 iterations. The means of the four upload speed clusters were 5.87 Mbps, 11.55 Mbps, 17.57 Mbps, and 38.62 Mbps. We observe that the upload cluster means obtained through the BST methodology are close to the actual offered upload speeds by ISP-A. BST achieves an accuracy of 99.3% for this set of upload speed measurements. This result validates our hypothesis and demonstrates the ability to use upload speed to narrow down potential subscription plans from which a given speed test may originate.

Download Speed Subscription Tiers. After determining the upload speed cluster of the speed test measurements, we apply the BST methodology within each of the four clusters of upload speed. Figure 5 shows the clusters of download speeds present within the upload speed clusters identified in the previous step.

Tier 1-3: This cluster consists of measurements from users subscribed to the 5 Mbps upload speed. Within this tier, we label the three available download speed plans as Tier 1, Tier 2, and Tier 3 to refer to the offered 25 Mbps, 100 Mbps, and 200 Mbps download speeds, respectively. Because the MBA dataset does not have the 25 Mbps download plan, our analysis consists of Tiers 2 and 3. There are 15, 781 measurements total from Tiers 2 and 3 in the MBA-State-A dataset. From Figure 5(a), we see two major download speed peaks after applying the KDE method to the download speeds in this cluster.

After determining the number of clusters, we apply the BST methodology to attach each download speed measurement point to the appropriate subscription tier class. The means of the two clusters found by BST are 110.89 Mbps and 231.69 Mbps, which are greater than the advertised download speeds. This observation indicates that ISP-A provides performance that surpasses the subscribed download speed for these subscription tiers. Previous studies [47] observed similar ISP behavior in the past. In comparing our calculated download speed plan with the ground truth, we determine that our methodology can accurately identify 100% of the download speed measurements in this cluster.

Tier 4: There are 4, 185 measurements in the MBA-State-A dataset that belong to this subscription cluster. The upload speed in this cluster is 10 Mbps; only one plan offers this upload speed, with a 400 Mbps download speed. Though our methodology achieves 100% accuracy in determining the subscription tier of these measurements, the KDE method reveals several download speed peaks within this cluster (see Figure 5(b)). We apply the BST methodology to detect four download speed clusters. The four means obtained through the process are 333.48 Mbps, 335.15 Mbps, 400.37 Mbps, and 463.31 Mbps. While it is unclear why four clusters are detected,



(a) Tier 1-3 **(b) Tier 4** **(c) Tier 5** **(d) Tier 6**

Figure 5: Download speed density using the KDE method within each cluster of upload speed. Black vertical lines represent the corresponding download speed plans for each upload speed.

it could be due to ISP throttling. It remains future work to diagnose the exact cause.

Tier 5: There are 2,453 measurements in this cluster, and the offered upload speed is 15 Mbps. This tier offers a download speed of 800 Mbps. Like Tier 4, BST achieves 100% accuracy in determining this subscription tier. Figure 5(c) shows a peak at around 700 Mbps, closer to offered speed, with the KDE method. We also observe multiple peaks around 300 Mbps and 400 Mbps. The BST methodology detects three clusters of download speed with means 269.98 Mbps, 358.06 Mbps, and 705.35 Mbps. We observe an overlap in download speed tier means between tiers 4 and 5. However, the proposed BST methodology isolates the download speeds into their respective subscription tiers.

Tier 6: ISP-A offers a plan with download speed 1200 Mbps and 35 Mbps upload speed; BST achieves 100% accuracy in inferring this subscription tier. In State-A, there are 3,508 measurements in this subscription tier. Figure 5(d) shows a single major cluster of download speed after applying the KDE method. The BST methodology computes the mean of this download speed cluster to be 892.05 Mbps. This mean value is much lower than the offered download speed for this subscription class. This result shows the limitation of speed test-like measurements in saturating the available bandwidth in the higher end of the offered subscription plans. Previous work [39] made similar observations.

These promising results indicate the ability to infer subscription tier information for crowdsourced speed tests. In the following sections, we use the BST methodology to contextualize Ookla and M-Lab speed test measurements with subscription tier information.

5 AUGMENTING OOKLA & M-LAB DATA

Now that we have demonstrated the accuracy of our BST methodology, our next step is to apply our approach to contextualize crowdsourced speed test measurements. This step is critical to the interpretation of speed test data; by comparing speed test results to the subscribed broadband plan, we can gain insight into whether the network is under-performing. Our analysis in this section focuses on City-A.

5.1 Contextualization with Subscription Plans

Upload Speed Subscription Tiers. The measurement nodes for the MBA project collect data directly from the cable modems, increasing the accuracy of capturing the access network performance.

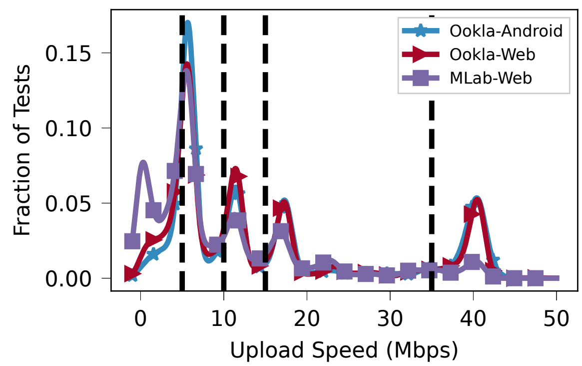


Figure 6: Upload speed density using the KDE method on City-A speed test measurements. The vertical lines represent the offered upload speed in each ISP-A plan.

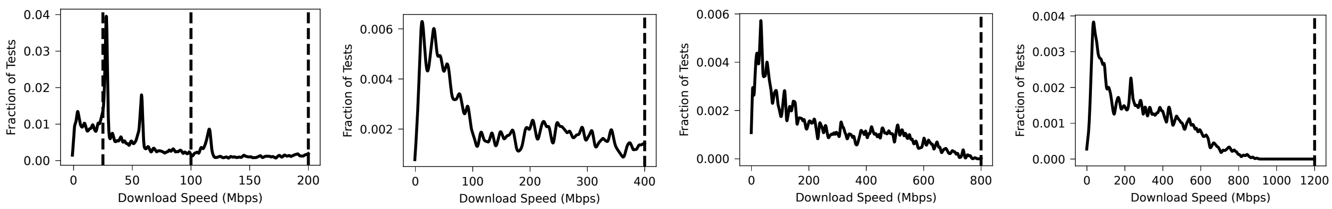
Unfortunately, a significant fraction of the speed test measurements in the Ookla and M-Lab datasets stem from end-user devices connected through a first hop WiFi link. The introduction of this single wireless link can significantly impact the speed test performance and can introduce additional skews [39, 44]. However, given the small range of possible maximum upload speeds, we hypothesize that it should still be possible to cluster these crowdsourced active measurements based on the recorded upload speed.

Figure 6 shows the upload speed densities for speed test takers who accessed Ookla using either the native Android application or the web-based portal, as well as M-Lab tests run through the web-based portal. Similar to the peaks in the MBA data shown in Figure 4, we observe densities of upload speed in the crowdsourced measurements that peak near the ISP-A offered upload speeds for all datasets. In addition to the four major peaks, there is an additional upload speed cluster in the 1 Mbps region in the M-Lab data.

We apply the BST methodology to associate the upload speed measurements to the four peaks around the ISP-A-provided upload speeds. Table 3 presents the number of measurements and means for the upload speed clusters (corresponding to an ISP subscription upload speed tier) detected by the BST methodology, broken down by device type when possible (Tables 5 – 7 in the appendix present the same breakdown for Cities B-D). We observe the means of each

Table 3: Number of measurements and the means (Mbps) for upload speed clusters that form near the ISP-A offered upload speeds in City-A. For each dataset, the means are obtained using the BST methodology.

Platform	Type	Tier 1-3		Tier 4		Tier 5		Tier 6	
		#Measurements	Mean	#Measurements	Mean	#Measurements	Mean	#Measurements	Mean
Ookla	Android-App	8,890	5.25	3,088	11.29	2,810	17.04	5,152	40.23
	iOS-App	33,265	5.30	13,299	11.35	9,530	16.71	19,480	39.82
	Desktop WiFi-App	4,551	5.54	1,377	11.59	3,638	16.82	1,750	39.92
	Desktop Ethernet-App	1,031	5.69	746	11.65	1,400	16.95	2,098	40.13
	Net-Web	43,833	5.72	12,802	11.64	29,157	16.69	15,797	40.06
M-Lab	NDT-Web	70,789	5.32	17,014	10.74	16,417	16.71	9,490	39.94



(a) Tier 1-3 (b) Tier 4 (c) Tier 5 (d) Tier 6

Figure 7: Download speed density using the KDE method within each upload speed cluster of Ookla Android device measurements.

cluster to be similar across all datasets. These means are also consistent with the means detected in the State-A dataset in section 4.3 for ISP-A offered plans. Given this similarity, we can associate the crowdsourced measurements to their subscription tier.

Download Speed Subscription Tiers. The much larger download speed plans offered by ISP-A and the performance variability caused by the end user's home wireless link create considerable challenges to clustering the measured download speeds. Figure 7 shows the densities of download speeds recorded by Ookla tests conducted on Android devices within each cluster of upload speed. The download speed densities for other Ookla device types and M-Lab are presented in Table 4 in the appendix.

There are five major download speed clusters in Tiers 1-3 of the Android dataset[4]. This number is three more than the number detected in the MBA State-A dataset for the same cluster and two more than what is offered by ISP-A for this subscription tier. After applying the BST methodology, we associate the download speed measurements to five clusters of download speed with means 8.04 Mbps, 27.14 Mbps, 57.85 Mbps, 115.65 Mbps and 214.01 Mbps. We associate the measurement points that belong to the components with mean values of 8.04 Mbps and 27.14 Mbps to Tier 1 as these measurements are close to the offered download speed. Similarly, we assign the measurements associated with clusters of mean values 57.85 Mbps and 115.65 Mbps to Tier 2. Finally, we associate measurements in the cluster of mean 214.01 Mbps to Tier 3.

Compared to the clusters formed by tests conducted over WiFi access links, the measurements in Tier 1-3 run by desktop devices connected with wired links (presented in Table 4 in appendix) produce three download speed clusters with means of 16.04 Mbps, 93.76 Mbps and 231.44 Mbps. These three means are closer to the three offered download speeds provided by ISP-A for this subscription tier.

We know that ISP-A offers a single download speed for each of the other upload speed tiers. However, Figure 7 indicates a large number of download speed clusters at various magnitudes. We apply the BST methodology and associate measurements with 10 clusters of download speed for each of tiers 4-6. Table 4 in the appendix presents the download speed cluster mean values that belong to each upload speed cluster. The number of components detected for wired measurements in each of these tiers is less than in wireless ones.

For Tier 4, we observe three clusters with mean values of 67.77 Mbps, 288.29 Mbps, and 461.18 Mbps. For Tier 5, we identify two groups with mean values of 146.46 Mbps and 595.59 Mbps. We also observe two clusters for Tier 6, with mean values of 103.96 Mbps and 906.87 Mbps. The wide range of values represented by these download speed clusters means, as well as for WiFi tests, indicates a significant variance in the results of the speed tests. This result further justifies our approach of first clustering these measurements using the less noisy, slower upload speeds before associating the measurements with complete subscription tier information.

WiFi-connected devices contribute to almost 97% of the native application tests in the Ookla dataset. Roughly half of these tests originate from the lowest subscription tier. As a result, if we take any aggregate (such as the median) of speed test data in a locality, we would, at best, get a representation of the Internet quality obtained by the lower subscription tiers, as opposed to the complete picture. Contextualizing these measurements with subscription tier

[4]All Android measurements occur over WiFi.

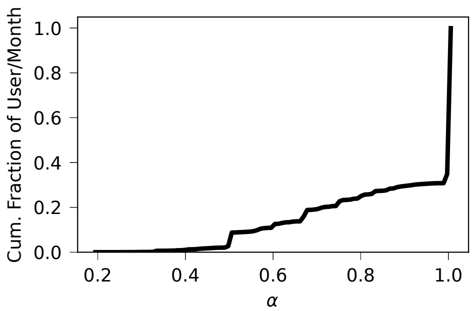


Figure 8: CDF of α values per user per month.

information is crucial before making any general assessment of the Internet quality in a region.

5.2 Investigation of Consistency

Because we lack ground truth for the Ookla and M-Lab measurements, we turn to other approaches to evaluate the accuracy of our BST methodology in these noisy environments. In this section, we analyze the consistency of BST in its association of speed test measurements with subscription tiers. To do so, we focus on users who conducted more than five speed tests in a month, and we examine whether each measurement from a single user is assigned to the same subscription plan, or whether there is variability in the assignment.

For every user u in month m, we determine the ratio r of tests that were associated with each of the six subscription tiers. For the i^{th} cluster, this can be denoted as:

$$r_{ium} = \frac{N_i}{\sum_{k=1}^{6} N_k} \tag{1}$$

where N_i is the number of tests associated with tier i. We denote α as the maximum of these four ratios to represent the tier that had the highest portion of tests associated for a given user in some month i.e $\alpha_{um} = \max_{i\in\{1,..,6\}} r_{ium}$. A higher α value indicates that our BST methodology is consistent for an user across multiple tests in a month. Conversely, if multiple tiers are associated for a user in a month, α will be lower. Figure 8 shows the distribution of the α values recorded for users during the 12 months in 2021. The skew of α values towards 1 indicates that, for most users in a month, our BST methodology associates the user to a single tier the majority of the time (the median is 1).

6 DIAGNOSING SPEED TEST PERFORMANCE

The association of subscription tier to speed test measurement provides the context needed to determine whether a measurement indicates under-performance relative to the purchased plan data rate. Armed with this information, our objective is now to determine the potential causes of speed test measurements failing to achieve performance close to their subscription plan upload and download speed maximums. For ease of presentation, we present the analysis in this section on measurements from City-A; we verify separately that our findings are consistent with the other three cities. Additionally, because tiers 1-3 for ISP-A in City-A all share the same upload speed, we combine these measurements into one group for the analysis in this section. Finally, we focus the majority of our presentation on download speed due to its greater variability and susceptibility to performance degradation.

6.1 Effect of Home Network and Device

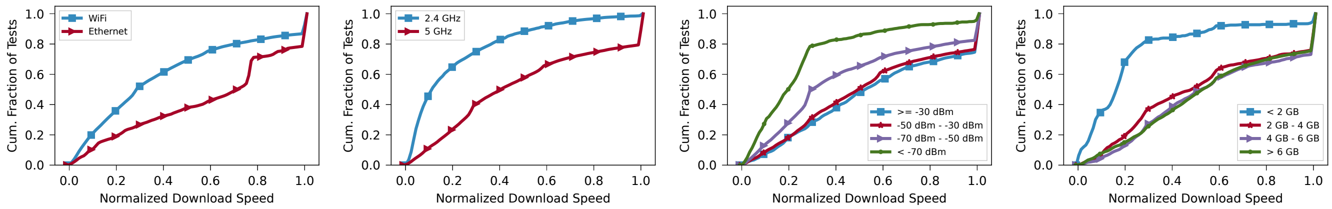
Previous work [39, 47] has documented that the home WiFi link can act as a significant barrier to saturating available bandwidth in the access network. Therefore, our first objective is to understand whether and how characteristics of the client's home network configuration lead to speed test under-performance with respect to the maximum bandwidth of the subscription plan. Our analysis in this section is possible because we can contextualize the measurements with their respective subscription tier information using our BST methodology. To capture any performance impacts, for every measurement, we normalize the recorded download speed by the offered download speed for the subscription tier. In the following, we quantify the number of speed tests that we identify as affected by different characteristics of the home WiFi link. We then study the effect of kernel memory limitations in the user device on speed test performance.

Access Link. Given the challenges and complexities of WiFi communication, our first step is to compare the speed test results that were conducted over WiFi with those from desktop computers connected to the home network via Ethernet. For this study, we include speed test measurements from all subscription tiers. We examine all WiFi speed tests conducted via Android, iOS and desktop devices (the Ookla-web and M-Lab datasets do not contain metadata about device/access type, and so these are not included in this analysis). Where relevant, we compare the WiFi performance of these devices with that of desktop devices connected through Ethernet.

As can be observed from Figure 9a, the difference in the normalized download speed distributions of WiFi and Ethernet access links is significant. For speed tests conducted over a WiFi network, the median normalized download speed is 0.28. This value is almost three times less than the median normalized download speed of 0.71 for Ethernet speed tests. We observe similar results for other cities. Without proper contextualization, the lower download speeds from tests conducted over WiFi could be misconstrued to be under-performance of the provider network.

WiFi Band. Next, we more deeply examine WiFi speed test performance and investigate the impact of the WiFi spectrum band on download speed. Modern routers are equipped to operate in both the 2.4 GHz and 5 GHz WiFi bands [36, 45]. The 5 GHz band supports greater bandwidth while more susceptible to attenuation compared to the 2.4 GHz band [33]. Amongst our datasets, only the Ookla Android measurements contain information about the WiFi band a device used during the speed test. About 23% (15k) of all Android measurements were conducted over the 2.4 GHz WiFi band; the remaining were on the 5 GHz band.

We normalize the reported download speed by the respective ISP offered download speed within a subscription tier. Figure 9b shows the distribution of the normalized download speed for all Android measurements separated by the WiFi band. The figure shows



(a) Access Type **(b) WiFi Band** **(c) RSSI** **(d) Available Memory**

Figure 9: Impact of WiFi characteristics and available memory on speed test performance.

a striking difference between the performance of tests in the two bands. While the median normalized download speed is just 0.11 for 2.4 GHz speed tests, it is 0.4 for 5 GHz tests. This median difference in performance between these two bands is amplified for higher subscription tiers. For Tier 6, the median normalized download speed for 5 GHz speed tests (0.25) is over six times more than that of 2.4 GHz measurements (0.04). This finding demonstrates that the WiFi spectrum utilization has an outsized impact on speed test performance, and again, without proper contextualization, the cause of the lower performance on 2.4 GHz devices could be misconstrued.

WiFi RSSI. We next analyse the impact of WiFi RSSI on speed test performance. As our analysis previously demonstrated, 2.4 GHz tests under-perform compared to 5 GHz tests. Hence, for this analysis we only consider the tests conducted in the 5 GHz WiFi band in the Ookla Android dataset. We bin the tests into four categories of WiFi RSSI values. Similar to the access type and WiFi band analysis, for each RSSI bin, we calculate the distance between the measured and subscribed performance for each test. Figure 9c shows the distribution of the normalized download speed achieved by speed tests for each RSSI bin.

9% of the 5 GHz Android tests have RSSI values lower than −70 dBm; these tests record the lowest median normalized download speed of 0.2. The median normalized download speed increases to 0.3 for the speed tests conducted in the WiFi RSSI region −70 dBm - −50 dBm; these tests account for 49% of 5 GHz Android speed tests. The next RSSI bin (−50 dBm - −30 dBm) contains 37% of the total 5 GHz Android speed tests; these tests recorded a median normalized download speed of 0.49. Finally, 5% of all 5 GHz Android speed tests had an RSSI better than −30 dBm; the median normalized download speed for these tests was 0.52. As shown in figure 9c, the performance difference varies by over a factor of two between the lowest and highest RSSI bins for all subscription tiers. This difference increases to more than five when considering speed tests in Tier 6. It is therefore critical to contextualize WiFi speed test measurements with signal strength as poorer RSSI can significantly affect the measured performance.

Kernel Memory. We next study the memory available to the Android device kernel during the speed tests to understand its role in achieved performance. For Android measurements, Ookla reports the amount of memory (in megabytes) available to the kernel. To minimize the impact of other factors, we only consider Android measurements in the 5 GHz WiFi band with an RSSI better than −50 dBm (9k measurements).

We bin the available kernel memory into four groups: less than 2 GB, 2 GB – 4 GB, 4 GB – 6 GB and more than 6 GB. Figure 9d presents the CDFs of the distance between subscribed and achieved speed test performance grouped by available kernel memory. The distance increases as less memory is available to the kernel during the speed test. 7% of measurements have less than 2 GB of available kernel memory. This group of measurements also recorded the smallest median normalized download speed of 0.16. The next two bins each contribute 17% of the speed tests. The median normalized download speed is 0.48 and 0.52 for 2 GB – 4 GB and 4 GB – 6 GB of available kernel memory, respectively. The majority of speed tests (59%) are issued from devices with over 6 GB of available memory; these tests record the highest median normalized download speed of 0.53, three times more than the 2 GB tests. This difference increases further for higher subscription tiers with Tier 6 tests recording a difference of five times in median normalized download speed between these two groups. This result shows that speed test performance can be greatly impacted by available memory and is therefore another important piece of context for speed test measurements.

Combination of Local Effects. In our final analysis of the impact of local characteristics on speed tests, we divide the entire Android dataset, across all subscription tiers, into two groups. The first group contains measurements that were conducted on 5 GHz WiFi band, with better than −50 dBm RSSI, and with more the 2 GB of available kernel memory. Based on our results in figure 9, this group of tests should experience the lowest impact of the home network and device characteristics on achieved speed test performance. We, therefore, term this group "Best".[5] Conversely, the measurements that do not belong to this group are placed in the "Local-bottleneck" group, as they are more likely to experience constraints from the home network or device memory. It is worth mentioning that the Ookla Android dataset does not provide metadata about other potential local impacts, such as WiFi interference and WiFi channel occupancy. In the absence of this information, we are restricted to the subset of local characteristics presented in this analysis.

In total, 61% (∼12k) of all Android measurements belong to the Local-bottleneck category. This indicates that the performance of the majority of speed tests is likely negatively impacted by home network or device characteristics. Figure 10 presents the normalized

[5] We do not claim that this group of tests does not have other bandwidth constraints, such as a poorly performing cable modem, or a faulty access link, etc. The labeling of "Best" reflects the fact that, amongst the context we investigate, this group of measurements is least likely to experience performance limitations.

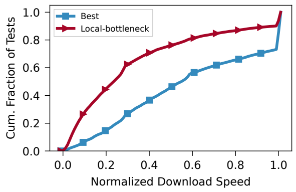


Figure 10: Comparison of normalized download speed with and without local bottlenecks.

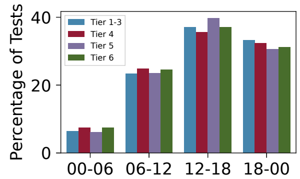
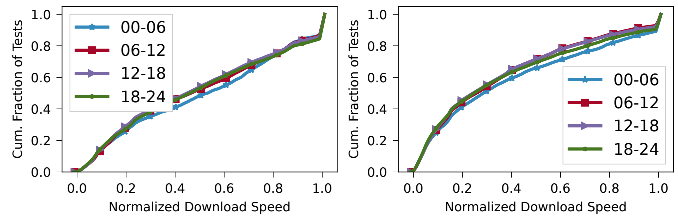


Figure 11: Percentage of speed tests in each six hour time bin.



(a) Tier 4 **(b) Tier 5**

Figure 12: Normalized download speed between measured and offered values for Ookla tests based on time of day.

(with respect to the respective subscription tier) median download speed recorded by both groups. The difference in performance is captured by the median normalized download speed of 0.22 for Local-bottleneck tests, over twice as low as the 0.52 achieved by “Best" tests.

6.2 Time of Day Effect

To study the effect of the time of day of a speed test, our first step is to determine the percentage of speed tests that originate at each time of day, for each subscription tier. With this data, we can then analyze the download speed performance, per tier, to determine whether there are measurable differences based on time of day. To explore this time of day effect, we bin the tests into four 6-hour periods: 12am-6am (00-06), 6am-12pm (06-12), 12pm-6pm (06-18) and 6pm-12am (18-00), all with respect to local time of the user. For each time bin, we calculate the percentage of speed tests issued by each subscription tier across all devices in the Ookla dataset; Figure 11 shows the result. We observe that there is not a significant difference in the percentage of speed tests in each time bin by subscription tier. We observe a similar trend across all subscription tiers in the M-Lab dataset, but omit these results for brevity. The smallest percentage of tests occur during the night and early morning hours, while the majority of tests, across all subscription tiers, occurs in the afternoon and evening/early night hours. This finding is contrary to the observation made in [46], where it was reported that speed tests are primarily issued during the day.

We next explore whether the performance measured by each speed test differs based on the time it is executed. In particular, our objective is to evaluate how much further (or closer) measurement download speeds are compared to the subscription plan maximums based on the time of day. With this approach, we will be able to quantify whether performance drops are more likely to occur during specific time periods.

Figure 12 shows the CDFs of the normalized download speed for two subscription tiers across all device types. Our results demonstrate that the speed test performance with respect to the subscribed performance remains similar across all time bins within the day, with slightly better performance recorded for tests conducted during 00-06 hours. For example, the median normalized download sped for iOS tests for Tier 4 are 0.53, 0.46, 0.45 and 0.46 during the 00-06, 06-12, 12-18 and 18-24 time periods, respectively. Similarly, when we analyse the results in the higher subscription tiers, we observe slightly better median normalized download speeds during the off-peak time periods (e.g. 00-06). The median distances for Tier 5 tests are 0.21, 0.19, 0.18 and 0.19 during the 00-06, 06-12, 12-18 and 18-24 time periods, respectively. Based on these results (and similar results for M-Lab data), we conclude that the time of the test does not play a meaningful role in the achieved performance.

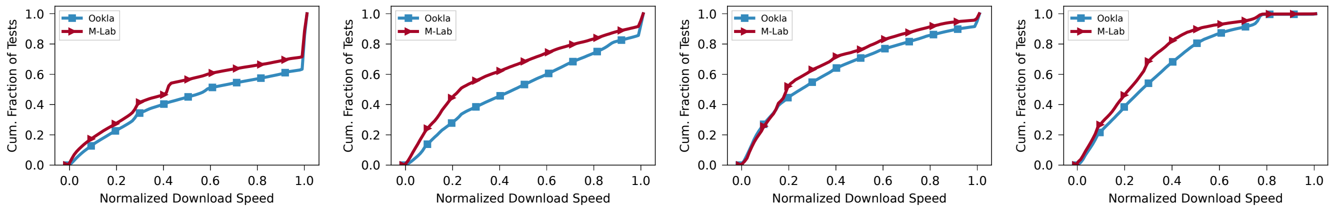
6.3 Effect of Speed Test Vendor

As stated earlier, Ookla and M-Lab are two of the most popular speed test vendors, and hence the datasets on which we base our study. However, there are some key methodological differences between their speed test measurements. Critically, M-Lab’s NDT conducts its speed test measurements with a single TCP thread, while Ookla speed tests utilize multiple threads [21, 39]. Prior work has found that, as a result, the M-Lab speed test suffers from underestimation of the available bandwidth [35, 37, 39]. In this section, our objective is to quantify the amount by which the performance reported by Ookla and M-Lab measurements differs. Because we have been able to associate speed test measurements with their subscription tiers, we have the ability to closely compare speed tests that, in theory, should achieve similar performance. Hence, in this study, we are able to compare Ookla and M-Lab measurements within the same subscription tier, for the same city, and the same ISP.

Figure 13 shows the distributions of the distances between subscribed and achieved performance for each subscription tier for Ookla and M-Lab measurements in City-A and ISP-A. Across all tiers, M-Lab measurements record greater distance from the subscribed performance than Ookla tests. For tiers 1-3, the median normalized download speed of M-Lab (0.83) is roughly 1.2 times worse than that of Ookla (1). Similarly, the factors by which M-Lab’s median normalized download speed lags Ookla’s are 2, 1.4 and 1.2 for tiers 4-6, respectively. As a result of these differences, it is critical for users of each test to understand what each test measures before drawing any specific conclusions, or making policy recommendations, based on performance results.

7 RELATED WORK

Multiple prior studies have characterized crowdsourced speed test measurements to better understand their utility and usability. In [44], the performance of three million Ookla measurements from 15 cities was analysed. The results demonstrated the high



(a) Tier 1-3 (b) Tier 4 (c) Tier 5 (d) Tier 6

Figure 13: Comparison of Ookla and M-Lab speed test normalized download speed per subscription tier.

variability that exists in speed test measurement, particularly for wireless tests. However, the work did not analyse the impact of any factors in impeding speed tests from achieving subscribed performance. The authors in [35] benchmarked Internet performance across multiple metro areas using Ookla speed tests. Their analysis reveals the presence of a large number of low performing speed tests in all cities. Recently, [37] studied the M-Lab dataset and highlighted the need for proper contextualization of measurements prior to drawing generalizable conclusions. The authors of [46] demonstrated the shortcomings of crowdsourced measurements in detecting overall Internet congestion. In [43], the location and income group biases of speed test origin are analysed. The work in [40] illustrates the shortcomings of speed tests in terms of not reaching subscribed speed through a sample of 50 tests from a single home. Similar to our finding, their result shows that upload speed has a small variance compared to download speed. In [32, 39], a detailed analysis of factors that can impact speed test performance is presented. In comparison to these and other similar studies, our work goes significantly further, in part by adding ISP subscription tier context to quantify how close (or far) current speed test results are from actual subscribed performance.

Other prior work has analysed how local network factors can create performance bottlenecks. Local factors are demonstrated to create a bottleneck to achieving download speeds greater than 20 Mbps in [47]. The negative impact of suboptimal WiFi parameters was studied in [33, 38]. The work in [45] demonstrated that factors such as RSSI significantly affect the overall measured latency. In [41], the Secure Digital Input Output bus sleep in smartphone was identified as a large contributor to overall latency. Our study finds that the vast majority of measurements experience bottlenecks by home network and device characteristics, resulting in significant performance underachievement compared to the theoretical maximum of the subscribed broadband plan.

8 CONCLUSION

In this work, we develop a novel BST methodology to augment crowdsourced speed test datasets with ISP subscription tier information. This critical context enables us to analyze and quantify the impact of a variety factors that can degrade speed test performance. The extensive impacts we uncover, which at times differentiate performance more than seven-fold, underlines the need for meaningful contextualization of crowdsourced speed test measurements prior to drawing generalizable conclusions about regional broadband access and quality. This is particularly important for policymakers prior to basing funding and investment decisions on this data. We also highlight the need for speed test platforms used to challenge provider coverage claims to ensure their test methodologies maximize link throughput. We believe that the need for accurate broadband mapping has never been greater, and that crowdsourced speed test measurement platforms will provide an invaluable part of the data needed to generate these maps. We hope that our work contributes to the advancement of this critical mapping effort. As such, we make the following recommendations.

Recommendations: As part of the Broadband DATA Act [3], the FCC has outlined and continues to refine a process for consumers to challenge fixed and mobile provider coverage claims. As part of the challenge process, consumers can submit speed test measurements taken from specified tools. Our work has identified critical metadata that we believe must accompany each measurement. It is possible to collect some of these metrics, such as access link, WiFi RSSI, etc., without user-level intervention. However, extracting all the recommended metadata for all end hosts might not be possible depending on their operating systems and browsers. Thus, the measurement platforms should collect as much contextual information as possible to better understand the speed test measurements. Though it is possible to infer the subscription plan, we recommend collecting this information from as many users as possible. Our recommendation is motivated by the observation that subscription plans play a critical role in assessing Internet quality in a region. Importantly, *we believe the context we recommend must be coupled to (i.e., publicly accessible with) measurement results as meta-data so that such measurements can be properly analyzed.*

Note that we do not claim our study to be an all-inclusive list of needed context. Other factors, such as the make and model of the cable modem or additional relevant home router information, are likely also essential. However, they are out of the scope of this study. Finally, *we encourage all speed test vendors who wish to create platforms for such coverage challenges to ensure that the speed test is constructed so that it maximizes the throughput of the measured path.* Designing such test methodologies, especially for high-speed access links, is non-trivial and requires further exploration [39].

ACKNOWLEDGMENTS

We wish to thank the anonymous IMC reviewers for their valuable feedback on the paper. We also thank Ookla and Zillow for sharing their datasets with us. This work was funded through National

Science Foundation Rapid Response Research (RAPID) award NSF-2033946.

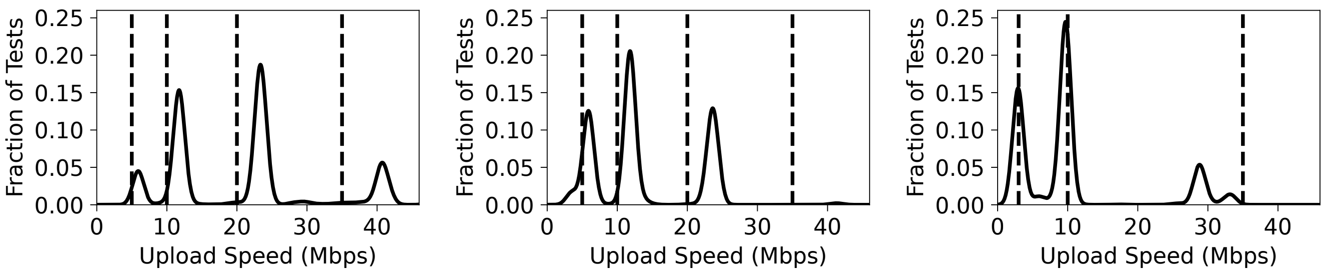
REFERENCES

[1] 2020. America's Racial Gap & Big Tech's Closing Window. Retrieved 10/05/2021 from https://www.db.com/newsroom_news/
[2] 2020. Expanding Internet Access Improves Health Outcomes. Retrieved 10/18/2021 from https://www.govtech.com/network/Expanding-Internet-Access-Improves-Health-Outcomes.html
[3] 2020. S.1822 - Broadband DATA Act. https://www.congress.gov/bill/116th-congress/senate-bill/1822.
[4] 2020. The Results Are In for Remote Learning: It Didn't Work. https://www.wsj.com/articles/schools-coronavirus-remote-learning-lockdown-tech-11591375078.
[5] 2022. Absent New FCC Broadband Maps, Local Govs Plot Coverage. https://www.govtech.com/network/absent-new-fcc-broadband-maps-local-govs-plot-coverage.
[6] 2022. Broadband Availability and Access. https://www.rural.pa.gov/publications/broadband.
[7] 2022. Broadband Speed: FCC Map Vs. Experience on the Ground. https://dailyyonder.com/broadband-speed-fcc-map-vs-experience-ground/2018/07/25/.
[8] 2022. Calspeed A Home Broadband Study. https://www.calspeed.net/index.html.
[9] 2022. Expectation–maximization algorithm. Retrieved 03/09/2022 from https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm
[10] 2022. FAST. https://fast.com/.
[11] 2022. Federal Communications Commission NOTICE OF PROPOSED RULEMAKING MB Docket No. 22-13. https://docs.fcc.gov/public/attachments/FCC-21-20A1.pdf.
[12] 2022. Fixed Broadband Deployment Data from FCC Form 477. https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477.
[13] 2022. Grow North encourages people to take internet speed test to help improve broadband infrastructure in the region. https://www.wxpr.org/business-economics/2022-03-29/grow-north-encourages-people-to-take-internet-speed-test-to-help-improve-broadband-infrastructure-in-the-region.
[14] 2022. H.R.3684 - Infrastructure Investment and Jobs Act. https://www.congress.gov/bill/117th-congress/house-bill/3684/text.
[15] 2022. Ingham County Asks Residents and Businesses to Participate in Survey on Broadband Internet Access and Speed. https://www.prnewswire.com/news-releases/ingham-county-asks-residents-and-businesses-to-participate-in-survey-on-broadband-internet-access-and-speed-301518936.html.
[16] 2022. Kernel density estimation. Retrieved 03/22/2022 from https://en.wikipedia.org/wiki/Kernel_density_estimation
[17] 2022. Measuring Broadband America. https://www.fcc.gov/general/measuring-broadband-america.
[18] 2022. Measuring Broadband Raw Data Releases - Fixed. https://www.fcc.gov/oet/mba/raw-data-releases.
[19] 2022. Millions of Americans can't get broadband because of a faulty FCC map. There's a fix. https://www.cnet.com/home/internet/features/millions-of-americans-cant-get-broadband-because-of-a-faulty-fcc-map-theres-a-fix/.
[20] 2022. Mixture model. Retrieved 03/08/2022 from https://en.wikipedia.org/wiki/Mixture_mode
[21] 2022. MLab Test Your Speed. https://speed.measurementlab.net/#/.
[22] 2022. SPEEDTEST. https://www.speedtest.net/.
[23] 2022. Speedtest by Ookla Global Fixed and Mobile Network Performance Map Tiles. https://github.com/teamookla/ookla-open-data.
[24] 2022. Speedtest is the world's #1 internet utility. https://www.ookla.com/analysis.
[25] 2022. The FCC's broadband map won't be ready for a year. This data company has already built one. https://www.cnet.com/home/internet/the-fccs-broadband-map-wont-be-ready-for-a-year-this-data-company-has-already-built-one/.
[26] 2022. The Speedtest Server Network™. Retrieved 08/22/2022 from https://www.ookla.com/network
[27] 2022. Unit Profile. http://data.fcc.gov/download/measuring-broadband-america/2021/unit-profile-sept2020.xlsx.
[28] 2022. USA, actively participate in FCC policy-making! https://samknows.com/blog/usa-actively-participate-in-FCC-policy-making.
[29] 2022. Welcome to SpeedSurvey: presented by The State of Alabama. https://www.google.com/search?q=ctc+alabama+speed+test&oq=ctc+alabama+speed+test&aqs=chrome..69i57j33i160.6405j0j4&sourceid=chrome&ie=UTF-8.
[30] 2022. Xfinity Speed Test. https://speedtest.xfinity.com/.
[31] 2022. Zillow's Transaction and Assessment Database (ZTRAX). Retrieved 03/22/2022 from https://www.zillow.com/research/ztrax/
[32] Steven Bauer, David D. Clark, and William Lehr. 2010. Understanding Broadband Speed Measurements. *Annual Research Conference on Communications, Information and Internet Policy* (aug 2010).
[33] Sanjit Biswas, John Bicket, Edmund Wong, Raluca Musaloiu-E, Apurv Bhartia, and Dan Aguayo. 2015. Large-Scale Measurements of Wireless Network Behavior. *SIGCOMM Comput. Commun. Rev.* 45, 4 (aug 2015).
[34] Anthony Brindisi. 2020. REPORT ON THE STATE OF BROADBAND ACCESS IN NEW YORK'S 22ND CONGRESSIONAL DISTRICT.
[35] Igor Canadi, Paul Barford, and Joel Sommers. 2012. Revisiting Broadband Performance. In *Proceedings of the 2012 Internet Measurement Conference (IMC '12)*.
[36] Germán Capdehourat, Federico Larroca, and Gastón Morales. 2020. A Nation-Wide Wi-Fi RSSI Dataset: Statistical Analysis and Resulting Insights. In *Proceedings of the 2020 IFIP Networking Conference (IFIP '20)*.
[37] David D. Clark and Sara Wedeman. 2021. Measurement, Meaning and Purpose: Exploring the M-Lab NDT Dataset. *Annual Research Conference on Communications, Information and Internet Policy* (aug 2021).
[38] Diego da Hora, Karel van Doorselaer, Koen van Oost, and Renata Teixeira. 2018. Predicting the effect of home Wi-Fi quality on QoE. In *Proceedings of the IEEE INFOCOM 2018 - IEEE Conference on Computer Communications (INFOCOM '18)*.
[39] Nick Feamster and Jason Livingood. 2020. Measuring Internet Speed: Current Challenges and Future Recommendations. *Commun. ACM* 63, 12 (nov 2020), 72–80.
[40] George S. Ford. 2021. Form 477, Speed-Tests, and the American Broadband User's Experience. *SSRN* (mar 2021).
[41] Weichao Li, Daoyuan Wu, Rocky K.C. Chang, and Ricky K.P. Mok. 2016. Demystifying and Puncturing the Inflated Delay in Smartphone-Based WiFi Network Measurement. In *Proceedings of the 12th International on Conference on Emerging Networking EXperiments and Technologies (CoNEXT '16)*.
[42] David Major, Ross Teixeira, and Jonathan Mayer. 2020. No WAN's Land: Mapping U.S. Broadband Coverage with Millions of Address Queries to ISPs. In *Proceedings of the 2020 Internet Measurement Conference (IMC '20)*.
[43] Udit Paul, Jiamo Liu, Vivek Adarsh, Mengyang Gu, Arpit Gupta, and Elizabeth Belding. 2021. Characterizing Performance Inequity Across U.S. Ookla Speedtest Users. https://arxiv.org/abs/2110.12038
[44] Joel Sommers and Paul Barford. 2012. Cell vs. WiFi: On the Performance of Metro Area Mobile Connections. In *Proceedings of the 2012 Internet Measurement Conference (IMC '12)*.
[45] Kaixin Sui, Mengyu Zhou, Dapeng Liu, Minghua Ma, Dan Pei, Youjian Zhao, Zimu Li, and Thomas Moscibroda. 2016. Characterizing and Improving WiFi Latency in Large-Scale Operational Networks. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '16)*.
[46] Srikanth Sundaresan, Xiaohong Deng, Yun Feng, Danny Lee, and Amogh Dhamdhere. 2017. Challenges in Inferring Internet Congestion Using Throughput Measurements. In *Proceedings of the 2017 Internet Measurement Conference (IMC '17)*.
[47] Srikanth Sundaresan, Nick Feamster, and Renata Teixeira. 2016. Home Network or Access Link? Locating Last-Mile Downstream Throughput Bottlenecks. In *Proceedings of the 2016 Passive and Active Measurement Conference (PAM '16)*.
[48] Xinlei Yang, Xianlong Wang, Zhenhua Li, Yunhao Liu, Feng Qian, Liangyi Gong, Rui Miao, and Tianyin Xu. 2021. Fast and Light Bandwidth Testing for Internet Users. In *Proceedings of the 18th USENIX Symposium on Networked Systems Design and Implementation (NSDI '21)*.

APPENDIX

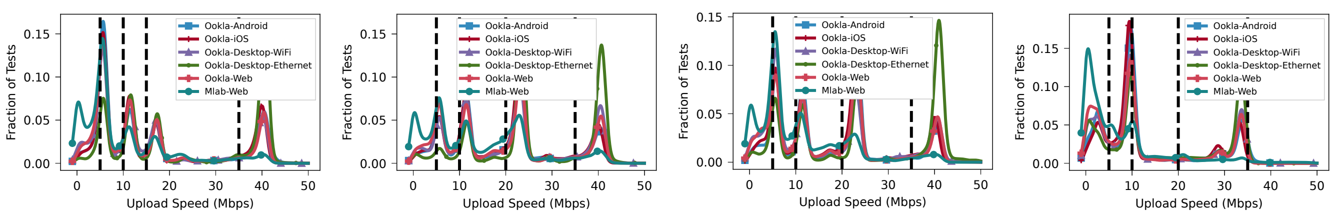
Table 4: Download speed means (Mbps) for each subscription tier in City-A. For each dataset, the means are obtained using the BST methodology.

Platform	Type	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5	Tier 6
Ookla	Android-App	8, 27	58, 116	214	21, 53, 93, 152, 212, 268, 327, 390, 445, 599	27, 73, 139, 219, 309, 403, 489, 574, 672, 879	40, 91, 160, 232, 304, 381, 461, 550, 636, 763
	iOS-App	9, 28	55, 84, 113	155, 197, 226	25, 57, 95, 144, 196, 244, 289, 337, 389, 442	28, 73, 121, 193, 264, 339, 421, 502, 589, 693	37, 88, 152, 223, 295, 367, 447, 535, 624, 737
	Desktop WiFi-App	15, 27	53, 86, 113	154, 202, 227	34, 77, 117, 155, 193, 251, 302, 340, 408, 453	22, 59, 105, 156, 211, 268, 345, 444, 540, 714	71, 177, 251, 345, 436, 540, 644, 735, 889, 1328
	Desktop Ethernet-App	16	94	231	68, 288, 461	147, 596	104, 907
	Net-Web	7, 28	55, 85, 114	170, 225	23, 55, 92, 146, 204, 265, 336, 405, 458, 637	19, 54, 97, 166, 239, 333, 437, 543, 692, 884	66, 162, 251, 350, 458, 568, 692, 820, 913, 1299
M-Lab	NDT-Web	6, 25, 47	100, 164, 221	18, 53, 84, 135, 196, 258, 337, 422, 569, 852	18, 53, 84, 135, 196, 258, 337, 422, 569, 852	22, 60, 105, 165, 229, 325, 413, 501, 652, 868	31, 93, 183, 260, 342, 429, 507, 610, 732, 892



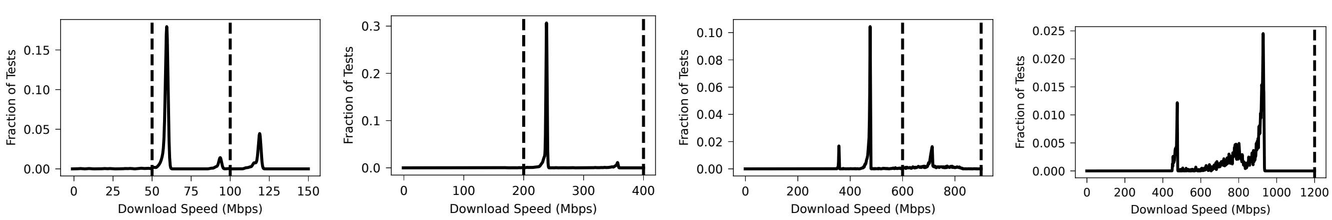
(a) State-B (b) State-C (c) State-D

Figure 14: Upload speed density using KDE method on MBA dataset for States B-D. The vertical lines are the upload speed plans offered by the dominant ISP in each state.



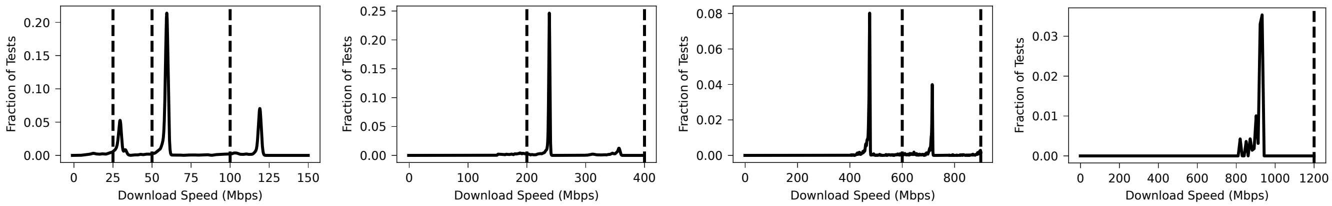
(a) City-A (b) City-B (c) City-C (d) City-D

Figure 15: Upload speed density using the KDE method on Cities A-D speed test measurements. The vertical lines represent the offered upload speed of the dominant ISP in each city.



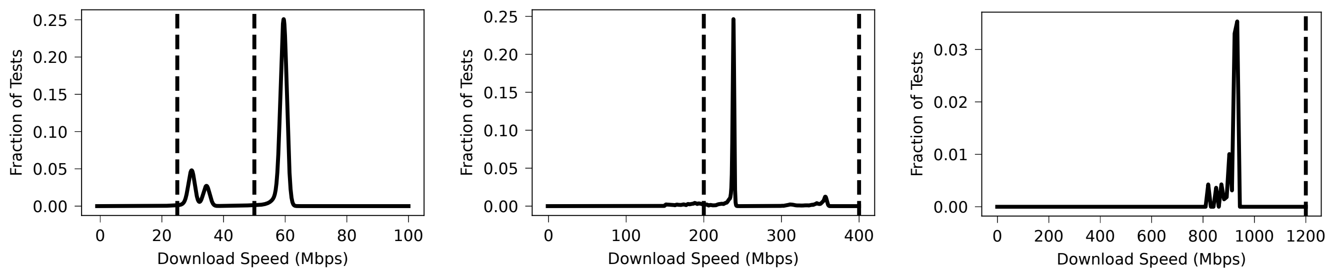
(a) Tier 1-2 (b) Tier 3 (c) Tier 4-5 (d) Tier 6

Figure 16: Download speed density using KDE method within each cluster of upload speed in State-B. Black vertical lines represent the corresponding download speed plans offered for each upload speed.



(a) Tier 1-3 (b) Tier 4-5 (c) Tier 6-7 (d) Tier 8

Figure 17: Download speed density using KDE method within each cluster of upload speed in State-C. Black vertical lines represent the corresponding download speed plans offered for each upload speed.



(a) Tier 1-2 (b) Tier 3-4 (c) Tier 5

Figure 18: Download speed density using KDE method within each cluster of upload speed in State-D. Black vertical lines represent the corresponding download speed plans offered for each upload speed.

Table 5: Number of measurements and the means (Mbps) for upload speed clusters that form near the ISP B offered upload speeds in City B. For each dataset, the means are obtained using the BST methodology.

Platform	Type	Tier 1-2 #Measurements	Tier 1-2 Mean	Tier 3 #Measurements	Tier 3 Mean	Tier 4-5 #Measurements	Tier 4-5 Mean	Tier 6 #Measurements	Tier 6 Mean
Ookla	Android-App	4965	5.73	2483	11.54	6794	22.42	2819	39.21
	iOS-App	18940	5.81	11358	11.48	29960	21.95	15042	38.08
	Desktop WiFi-App	2012	5.1	1281	11.48	3009	21.97	2093	39.01
	Desktop Ethernet-App	492	5.63	811	11.39	2048	23.32	2904	36.87
	Net-Web	30132	5.38	11925	11.56	37553	22.37	17504	39.62
Mlab	NDT-Web	144345	5.44	63805	11.16	135897	22.04	25553	39.23

Table 6: Number of measurements and the means (Mbps) for upload speed clusters that form near the ISP C offered upload speeds in City C. For each dataset, the means are obtained using the BST methodology.

Platform	Type	Tier 1-3 #Measurements	Tier 1-3 Mean	Tier 4-5 #Measurements	Tier 4-5 Mean	Tier 6-7 #Measurements	Tier 6-7 Mean	Tier 8 #Measurements	Tier 8 Mean
Ookla	Android-App	6766	5.28	3168	11.53	8307	22.28	3030	39.49
	iOS-App	11725	5.18	4711	11.45	12322	21.96	5579	38.84
	Desktop WiFi-App	2015	4.86	606	11.47	1094	21.61	854	38.21
	Desktop Ethernet-App	1020	4.92	628	11.48	868	23.36	2416	37.71
	Net-Web	24148	4.89	7982	11.54	21478	22.02	9697	39.53
Mlab	NDT-Web	34523	4.76	12789	10.72	13041	19.82	4416	35.47

Table 7: Number of measurements and the means (Mbps) for upload speed clusters that form near the ISP D offered upload speeds in City D. For each dataset, the means are obtained using the BST methodology.

Platform	Type	Tier 1-2 #Measurements	Tier 1-2 Mean	Tier 3-4 #Measurements	Tier 3-4 Mean	Tier 5 #Measurements	Tier 5 Mean
Ookla	Android-App	7244	3.51	8142	9.73	6462	28.69
	iOS-App	18598	3.72	26699	9.39	19177	28.03
	Desktop WiFi-App	2525	3.04	2233	9.59	2348	28.72
	Desktop Ethernet-App	1845	3.6	1716	9.68	3096	28.99
	Net-Web	40452	3.05	29642	9.7	27517	28.51
Mlab	NDT-Web	71833	2.95	61435	7.6	24541	24.94