You wanna see some real speed? Comparative Analysis of global Measurement Lab and Cloudflare speed test results

Friedemann Lipphardt
Max Planck Institute for Informatics
Saarbruecken, Germany
frlippha@mpi-inf.mpg.de

Malte Tashiro
IIJ Research Laboratory / SOKENDAI
Tokyo, Japan
malte@iij.ad.jp

Romain Fontugne
IIJ Research Laboratory
Tokyo, Japan
romain@iij.ad.jp

Abstract

Measuring the quality and throughput of an end user's internet connection is a non-trivial task that various providers such as Measurement Lab and Cloudflare have tried to solve. In this paper, we compare publicly available aggregated data from these two sources, examine their differences in measuring algorithms, look at their results on a global scale, and discuss the differences. We find that Measurement Lab reports higher values than Cloudflare in measured throughput comparisons by roughly 5 to 10% on a global scale. Measurement Lab tests also outnumber Cloudflare by a factor of 10, making Measurement Lab the vastly more popular testing platform. Furthermore, we examine the impact of a new server location on the speed test throughput results.

CCS Concepts

• **Networks** → **Network measurement**; **Network performance analysis**.

Keywords

internet measurements, speedtests, visualization

ACM Reference Format:
Friedemann Lipphardt, Malte Tashiro, and Romain Fontugne. 2025. You wanna see some real speed? Comparative Analysis of global Measurement Lab and Cloudflare speed test results. In *20th Asian Internet Engineering Conference (AINTEC '25), November 25–27, 2025, Manila, Philippines.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3763400.3763444

1 Introduction

The throughput of a broadband connection, colloquially known as "internet speed", is a frequently used metric when comparing internet connectivity. Internet service providers (ISPs) use this metric to advertise services, customers select contracts based on it, and policy makers assess regional network performance with it [7, 21]. Billions of dollars in public broadband funding are allocated based on speed test measurements. Furthermore, researchers use speed test data as a common metric in measurement papers [1, 5, 26]. Consequently, internet speed has become an important value for consumers, governments, and researchers.

AINTEC '25, Manila, Philippines

ACM ISBN 979-8-4007-1846-5/25/11
https://doi.org/10.1145/3763400.3763444

However, measuring internet speed is non-trivial as illustrated by various approaches and measurement providers available. Typically, a user performs a speed test by connecting to a measurement server and transferring data as fast as the network allows. Although these tests are simple to execute, different measurement tools may yield different results due to varying measurement algorithms, infrastructure, and routing policies [16]. Thus, one needs to be careful when drawing conclusions from these datasets.

We compare one year of throughput results from two prominent speed test providers, Measurement Lab (M-Lab) and Cloudflare, to raise awareness about dataset differences and provide usage guidance.

Our analysis compares topological and geographical coverage (Section 5), demonstrates data sanitization needs (Section 5.4), documents country-wise differences (Section 6.2), analyzes overall differences (Section 6.3), and studies server deployment impact (Section 6.4). This study shows that not only do these data sets highly differ in the number of results and networks covered, but also in the absolute throughput reported for the same networks.

2 Background

Internet speed is an important metric for both ISPs and users. ISPs advertise theoretical maximum throughput which may be difficult to achieve in practice. Users rely on speed tests to measure actual internet speed. Speed tests measure the time to download and upload data to a remote server. Results depend on several factors: theoretical throughput limits, network conditions (congestion, packet loss), measurement algorithms, number of connections, and server location.

Three prominent speed test providers are M-Lab, Cloudflare, and Ookla. M-Lab and Cloudflare give open access to their data (discussed in Section 4), Ookla only does so upon request through their "Ookla for good" program [24]. Unfortunately, we were not able to get access to Ookla data, which is why we focus on M-Lab and Cloudflare in this paper. Brief descriptions of the M-Lab and Cloudflare speed tests are provided below.

2.1 M-Lab

M-Lab [13] is a research institution primarily sponsored by Google and has conducted speed tests since 2009. Their test is available via a dedicated client application, a website, or by searching for "speed test" on Google, which integrated the test into the search result page.

M-Lab's speed test data is split into three sections: web100, ndt5 and ndt7. web100 is deprecated since 2019 and its data has not been used in this paper. We include both NDT5 and NDT7 tests and do not make a distinction between them, as NDT7 is a replacement

protocol, intended to supersede NDT5, not a competitor, and has since been adapted by the majority of tests [17]. If technically feasible, all future tests will be performed via NDT7.

NDT5 is an earlier version of the Network Diagnostic Tool using a custom TCP-based protocol and requiring Java applets or dedicated clients, causing accessibility issues due to Java applet deprecation and firewall traversal problems.

NDT7 is the updated version designed to address NDT5's limitations. It utilizes WebSockets over TLS enabling secure browser-based communication. NDT7 supports advanced TCP features for more accurate network performance assessment and improves browser compatibility and firewall traversal.

Both NDT5 and NDT7 perform their throughput test using a single TCP connection.

2.2 Cloudflare

Cloudflare is a CDN company providing various services. Their speed test reports throughput, latency, and jitter [6]. The test is available via website and unofficial client applications [11]. Cloudflare's speed test operates over HTTPS, employing HTTP/2 and HTTP/3 (QUIC) protocols. It measures download and upload speeds using multiple concurrent connections and evaluates latency and jitter by measuring response times.

2.3 Methodological Differences

The systematic differences observed stem from fundamental architectural differences between platforms:

Connection Architecture: M-Lab NDT uses a single TCP connection measuring "bulk transport capacity" [14], while Cloudflare uses multiple concurrent HTTP/2 and HTTP/3 connections [6], better reflecting modern web browsing behavior.

Infrastructure: M-Lab uses research-oriented servers at IXPs prioritizing measurement neutrality, while Cloudflare leverages its global CDN network optimized for content delivery and low latency.

Performance Impact: The 5-10% systematic difference we see in Section 6.3 reflects trade-offs between methodologies - M-Lab's single-stream approach provides conservative estimates while Cloudflare's multi-connection approach achieves higher aggregate throughput on high-latency links.

2.4 Ethics

This paper uses publicly available data, anonymized and consolidated by Measurement Lab [18] and Cloudflare [6]. This data is available for anyone, only requiring a mailing list registration. No additional personal data is used. Therefore, this paper does not raise any ethical concerns.

3 Related Work

Various studies focus on speed test services and their impact. Morton [19] examines factors affecting measurement results, identifying key variables that influence throughput measurements across different testing scenarios. Feamster et al. [10] discuss current challenges and recommendations. MacMillan et al. [16] compare Ookla and M-Lab speeds using household probes, revealing similar findings to ours: statistically significant but low differences (5-10%, cf. Section 6.3). Bajpai et al. [2] compare platforms with global probes and standardization efforts. Deng [9] analyzes M-Lab performance characteristics, while Bauer et al. [3] explain speed test structure and results.

4 Data Sources

We use publicly available speed test results from M-Lab [18] and Cloudflare [6], both hosted at the same BigQuery instance [12] due to their recent collaboration [22]. The BigQuery dataset provides structured access to millions of speed test results, enabling large-scale analysis across global networks. We focus on throughput data, as Cloudflare's AIM scores [25] cannot be calculated for M-Lab data.

We use weekly aggregations of mean download and upload throughput values, grouped by city and autonomous system number (ASN). Each record represents the average performance of all tests from a specific city-ASN combination during a given week. M-Lab uses Maxmind for geolocating clients [23]. Cloudflare reports server locations using airport codes (IATA), but their client geolocation method is unknown. Geographic precision varies between providers, potentially introducing systematic biases. We use one entire year (2023) of data from both sources, obtained from the M-Lab BigQuery instance.

5 Quantitative Comparisons

Before analyzing throughput values, we compare the topological and geographical coverage of both data sets.

5.1 Network coverage

Table 1: Network coverage comparison. Cell values indicate the number of overlapping ASNs of two datasets, whereas percentages (rounded) refer to the overlap with the set of total unique ASNs.

	M-Lab	CF	BGP	APNIC
M-Lab	40,058 (47%)	13,406 (16%)	39,046 (46%)	23,248 (27%)
Cloudflare	13,406 (16%)	14,371 (17%)	14,151 (17%)	10,616 (12%)
BGP	39,046 (46%)	14,151 (17%)	84,128 (99%)	25,054 (29%)
APNIC	23,248 (27%)	10,616 (12%)	25,054 (29%)	25,242 (30%)
All Four				10,310
Total				85,208

To understand the topological coverage of both datasets, we extract the ASes from where tests were launched and compare M-Lab and Cloudflare with two reference sets: (1) ASes originating IP prefixes on BGP and (2) ASes present in the APNIC population estimate. Both reference sets are based on a snapshot at the end of 2023. The BGP reference set is obtained using data collected by all RIS and Routeviews route collectors, representing the "reachable" Internet, whereas the APNIC set focuses on ASes with active users.

Table 1 shows the overlap between the different sets. The percentage value behind each entry refers to the overlap with the union of *all* sets, i.e., a total of 85.2k unique ASNs. We found that the set of originating ASes covers 98.7% of this superset, meaning that over 1% of ASes tested in 2023 have since stopped announcing prefixes on BGP.

For M-Lab we observe tests from 40k ASes, which represent a bit less than half (47%) of all ASes. Furthermore, M-Lab tests cover 23.2k (92%) of the APNIC set, resulting in excellent coverage of populated ASes. The Cloudflare dataset is significantly smaller with only 14.3k

ASes, even though 10.6k (74%) have inferred active users.M-Lab covers 2.8 times more ASes than Cloudflare, and comparing the coverage of the two speed test datasets we find that M-Lab covers 93% of the ASes appearing in the Cloudflare dataset.

Key takeaways: Even combined both datasets fail to cover the entire Internet, but show good coverage of ASes with active users. Furthermore, M-Lab is preferable in terms of raw network coverage.

5.2 City coverage

As we will discuss in Section 6.1, we compare the datasets based on their common factor: throughput values per city-ASN pair. In total, we observe almost 130k unique cities over both datasets. When paired with the 41k ASNs we observe a total of 648.6k unique city-ASN pairs, highlighting that most networks are spread across multiple cities, and that there can be numerous networks per city.

There is two key observations for the city-ASN pairs: (1) we find that M-Lab has a consistently higher coverage than Cloudflare, by two orders of magnitude, and (2) we see that both data sets contain new city-ASN pairs every week, which makes selection of a suitable timeframe for analysis critical.

To illustrate the number of city-ASN pairs contained in the data, we show the results for the first and last three weeks of our datasets in Table 2. We observe that the number of pairs stays stable over time and also that M-Lab has a significantly larger city coverage than Cloudflare. However, there are still some pairs unique to Cloudflare, showing that Cloudflare is nonetheless providing additional data.

Table 2: Weekly city-ASN pair counts for the first and last three weeks of 2023.

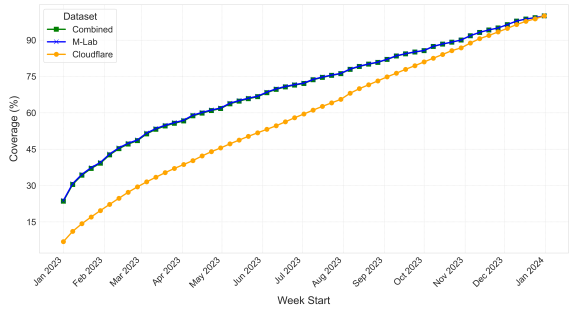
Week	M-Lab Only	CF Only	Both	Total
2023-01-01	147,358 (96.32%)	670 (0.44%)	4,964 (3.24%)	152,992
2023-01-08	158,370 (95.89%)	966 (0.58%)	5,829 (3.53%)	165,165
2023-01-15	154,675 (95.82%)	961 (0.60%)	5,782 (3.58%)	161,418
2023-12-17	146,631 (93.72%)	1,443 (0.92%)	8,385 (5.36%)	156,459
2023-12-24	141,677 (94.27%)	1,196 (0.80%)	7,415 (4.93%)	150,288
2023-12-31	148,796 (94.09%)	1,344 (0.85%)	8,009 (5.06%)	158,149

Now that we established how many city-ASN pairs we see each week, we check if there is an overlap between weeks.

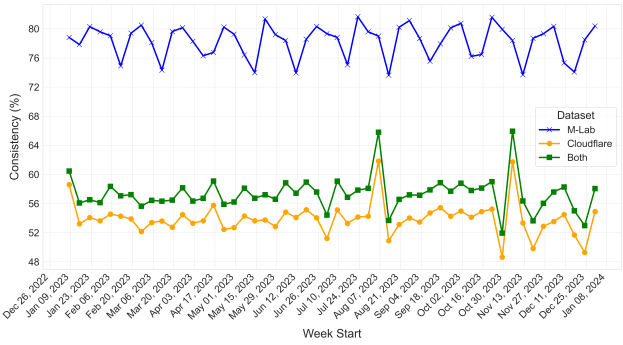
If we observe the same pairs every week, a single week of data would suffice for a comparative analysis.

First, we identify the time frame required to cover all city-ASN pairs seen during the year 2023 in each data set. Figure 1a shows the evolution of coverage if we start at the beginning of the year and add pairs every week. The "Combined" line, which reflects the combination of both data sets, follows the M-Lab line as the number of new pairs introduced by Cloudflare is low. Another observation is that there is a steady increase and we only see full coverage at the end of the year, i.e., during each week a consistent amount of new city-ASN pairs is uncovered. Therefore, a longer time frame always includes more city-ASN pairs.

We now measure the stability of the weekly pairs by comparing the overlap of pairs for adjacent weeks. While we saw above that both data sets introduce new pairs each week, Figure 1b shows that the weekly sets for M-Lab are more stable than Cloudflare. Between 75% and 85% of pairs overlap during adjacent weeks, whereas for Cloudflare the overlap is significantly lower with 49% to 62%. This



(a) The number of unique city-ASN pairs seen within the data is increasing steadily throughout the year 2023.



(b) Considering bi-weekly trends, the same city-ASN pairs are found within the M-Lab dataset at a much higher rate.

Figure 1: City-ASN Pair Trends

indicates that more users consistently run tests with M-Lab, making it better suited for a longitudinal study.

Key takeaways: For both data sets, a longer timeframe yields more coverage. Cloudflare data is less stable, indicating that the start and length of the time frame should be chosen with care.

5.3 IP version deployment differences

Metric	M-Lab	Cloudflare
IPv4 Tests	87,879,626	369,797
IPv6 Tests	28,017,070	112,826
IPv4 City-ASN Combos	345,140	38,459
IPv6 City-ASN Combos	73,765	13,519
Shared City-ASN Combos	42,958	6,367
V4-Only City-ASN Combos	302,182	32,092
V6-Only City-ASN Combos	30,807	7,152

Table 3: IP version comparison of NDT and Cloudflare test data, January 2023.

Using January 2023 data from both platforms (Table 3), IPv4 remains the dominant IP version in terms of test volume and deployment footprint:

- **Test volumes:** IPv4 tests account for roughly 75% of all measurements on each platform (M-Lab: 87.9 million IPv4 vs. 28.0 million IPv6; Cloudflare: 370 thousand IPv4 vs. 113 thousand IPv6).

- **Coverage breadth:** IPv6 city–ASN combinations represent only about 21%–24% of the IPv4 tally (M-Lab: 73.8 thousand vs. 345.1 thousand; Cloudflare: 13.5 thousand vs. 38.5 thousand).
- **Dual-stack reach:** Fully dual-stack (shared) city–ASN pairs make up a small fraction of the total—just 12% of IPv4 city–ASN combos in M-Lab (42.9 thousand of 345.1 thousand) and 17% in Cloudflare (6.4 thousand of 38.5 thousand).

This gap emphasizes that although IPv6 deployment continues to grow, IPv4 still commands both the majority of test traffic and the widest geographic/ASN footprint. We observe an early performance edge for IPv6 over IPv4, but detailed throughput analyses are provided later in this report (see Section 6.6).

Key takeaways: IPv4 tests outnumber IPv6 by roughly 3× on both M-Lab and Cloudflare; IPv6 coverage—measured by unique city–ASN pairs and dual-stack presence—remains a minority share of IPv4.

5.4 Data sanitization

Before comparing throughput values from the two data sets, we need to sanitize these values. As mentioned before, we aggregate speed test results per city-ASN pair over a week and use the average throughput values for comparison. However, some results are incomplete, usually missing the upload part of the test. In addition, many city-ASN pairs produce only a single result during the week. To draw meaningful conclusions we only include a city-ASN pair if there were at least three results during the week.

Table 4: Differences in data after cleaning and filtering.

Month		Initial Results	Cleaned (Rem. %)	Filtered (Rem. %)
2023-01	M	1,253,524	1,112,830 (88.78%)	797,901 (63.65%)
	C	90,624	81,660 (90.11%)	33,206 (36.64%)
2023-02	M	1,016,634	904,407 (88.96%)	646,148 (63.56%)
	C	80,118	74,894 (93.48%)	30,843 (38.50%)
2023-03	M	1,009,275	898,301 (89.00%)	634,296 (62.85%)
	C	80,299	74,749 (93.09%)	31,119 (38.75%)

M = M-Lab, C = Cloudflare

This threshold was chosen as a tradeoff between including as much data as possible and excluding data points that are not representative due to a small number of samples. Figure 2 shows the distribution of weekly occurrences per city-ASN pair for January 2023. There is an unstable tail of pairs with less than three results, which we cut off while retaining most of the datasets.

We summarize the impact of these steps in Table 4, exemplified with three months of data. The remaining months are similar, but not shown for brevity. As seen before, M-Lab has more results than Cloudflare by a factor of around 10. Excluding incomplete test results removes, on average, 11% of results from M-Lab and 6.5% from Cloudflare ("Cleaned" column). Finally, after filtering results with too few tests (under 3 as explained above) we are left with 62% of results for M-Lab and only 38% for Cloudflare. This sanitization increased the size gap between the data sets to a factor of 15, but is deemed necessary for a reasonable comparison.

Key takeaways: Both data sets need to be sanitized before analyzing them. M-Lab contains a larger fraction of incomplete results. Cloudflare results are less frequent and should be filtered for a relevant comparison. Due to the data set size, choosing the proper aggregation method is an important consideration.

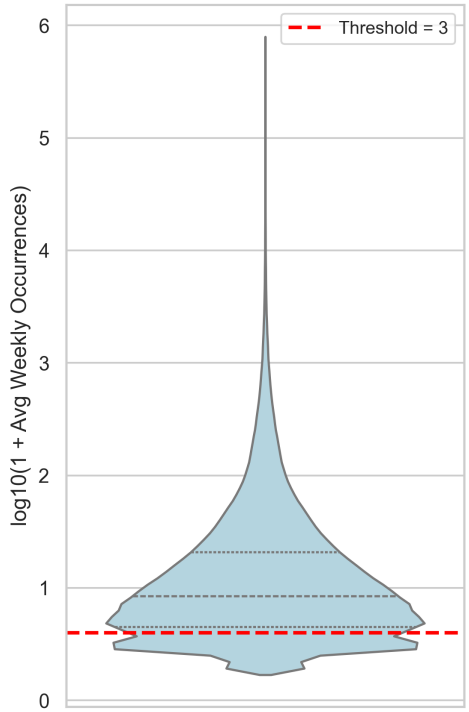


Figure 2: The distribution of average weekly occurrences per city-ASN pair for January 2023.

6 Throughput Comparison

6.1 Matching data points

Due to differences in available data, matching them for direct comparison is not straightforward. M-Lab provides more technical information than Cloudflare, so we use the largest common denominator: throughput values (download and upload) reported by both data sets.

To compare the data sets we group the results by a combination of the ASN and city of the client, followed by an aggregation in weekly intervals. Thus, we retain one result per city-ASN pair per week, representing the mean throughput of the tests run from this pair during the week. A more fine-grained grouping, e.g., by IP subnet or similar, would have been preferable but due to data limitations owing to privacy concerns by the two services, this was not possible. Again, as seen in other sections of this paper, the limiting factor is the Cloudflare data set, since M-Lab provides more specific data.

6.2 Outliers

Our first analysis looks at measurement outliers. We define the top outliers as the countries with the largest differences between mean upload and download speeds of the two data sets. We calculate the upload and download differences separately, and plot them on a scatter plot (not shown, but for an exemplary view of this type of scatter plot see Figure 4a), then measuring the absolute value of the distances from the origin. This way, we receive a comparable value for the differences between mean upload and download speeds

across all countries, and consider the ten countries with the largest differences as our top outliers, shown in Figure 3. By utilizing this method, we account for the difference in amount of tests per country by using mean values, aggregated on a yearly basis. Furthermore, this method uses cleaned data and we do not consider countries with less than 10 tests over the entire data period to be eligible.

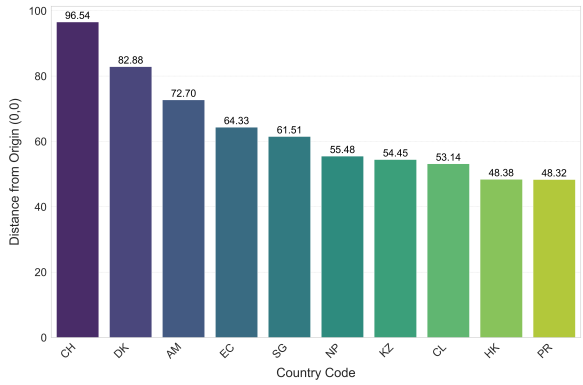
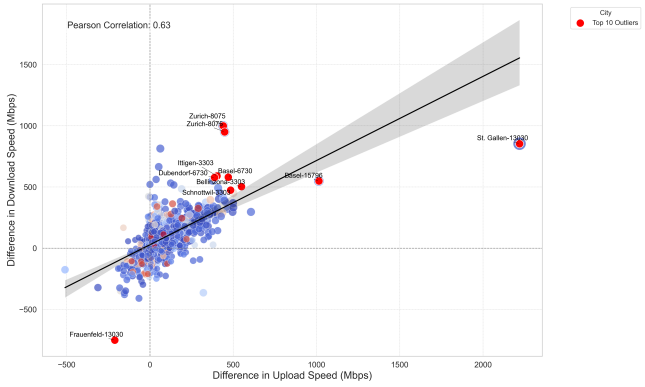
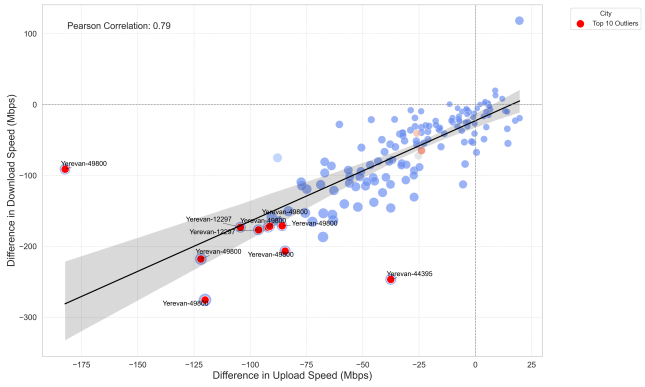


Figure 3: This bar plot for 2023 showcases the countries with the largest outlier score, which counts the number of total outliers and divides by unique ASNs in that country to account for an overrepresentation.



(a) Switzerland



(b) Armenia

Figure 4: Differences between upload and download throughput based on weekly data aggregation. Each point represents a city-ASN pair. Positive values mean that M-Lab throughput is higher and negative values mean that Cloudflare throughput is higher.

Using the method described above, we see that the top outliers for 2023 are Switzerland, Denmark, and Armenia. Looking at detailed results for these countries, we found that the outliers are caused by different factors. For Switzerland and Denmark (CH and DK), overall we observe higher download and upload throughput for M-Lab including a few city-ASN pairs with abnormally high throughput. For example in Figure 4a, we observe abnormally high throughput (up to 2 Gbps higher than Cloudflare) from Microsoft (AS8075) and Init7 (AS13030) around Zurich where the M-Lab server is located (hosted by Google).

For Armenia, however, Cloudflare is consistently faster than M-Lab across all city-ASN pairs (see Figure 4b). The server locations of both providers reveal that M-Lab has no server in Armenia and any of its surrounding countries, whereas Cloudflare is present in Armenia and almost all surrounding countries (Iran being the only exception).

Key takeaways: Using a refined calculation to account for differences in test quantity, we discover a number of countries to be outliers on a global comparison: Switzerland, Denmark and Armenia. This is caused largely by significant presence (or lack thereof) of server infrastructure in these countries by one of the testing providers.

6.3 Global

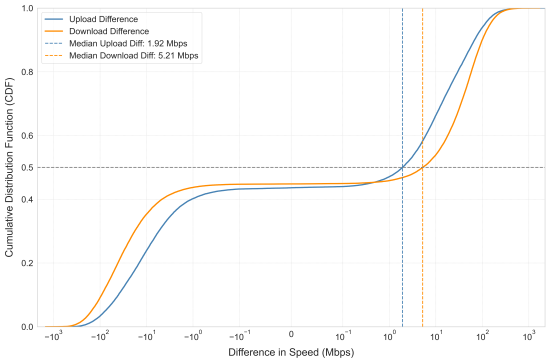
On the global scale, we see that the M-Lab throughput is usually higher than Cloudflare's throughput, but only by a small margin. Figure 5a depicts the throughput difference for all city-ASN pairs in 2023, the median value ($y = 0.5$) corresponds roughly to 5 Mbps speed difference for downloads and 2 Mbps difference for uploads. This 5-10% systematic difference represents a consistent bias across measurement platforms rather than random variation. In addition, the CDF reveals that uploads have a smaller variance, meaning that reported upload throughputs are more consistent. We posit that this is likely due to concurrent downloads on the server or client side.

Figure5b shows what we see in Figure 5a, but in a more detailed manner. The clear trend of the scatter plot towards the top right quadrant displays M-Lab outperforming Cloudflare in many of our observed city-ASN pairs. However, despite some outliers showing a drastic speed difference of up to 3000 Mbps, we note that this is an aggregation of differences between the mean upload and download differences over an entire year. If a city-ASN pair only runs both tests once, performing well on one but not the other, they will show up as an outlier in this plot. This plot is based on the cleaned, but unfiltered data set, therefore not accounting for minimum number of tests run.

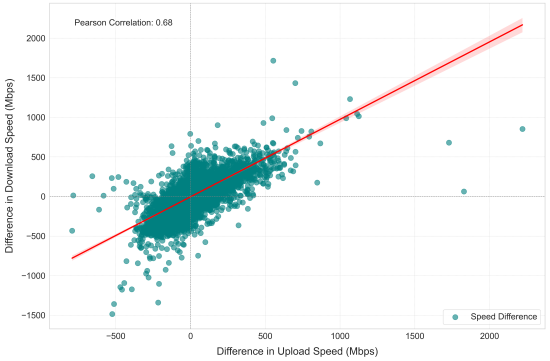
Key takeaways: M-Lab results consistently showcase a higher download throughput of roughly 5 Mbps and upload throughput of roughly 2 Mbps.

6.4 Servers

Another angle we consider is the impact of the server used to perform the speed test. In the context of this paper, when referring to server we are technically speaking about a server location, referred to by city or airport code in the data set. Both Cloudflare and M-Lab attempt to use the server geographically closest to the user [4, 8],



(a) While M-Lab consistently outperforms Cloudflare, the difference is marginal and the number of tests with significant differences low.



(b) While there are few outliers with a large speed difference, the vast majority of tests are found within the 4 center quadrants, with a statistically significant bias showing M-Lab outperforming Cloudflare.

Figure 5: Comparison of global data for 2023.

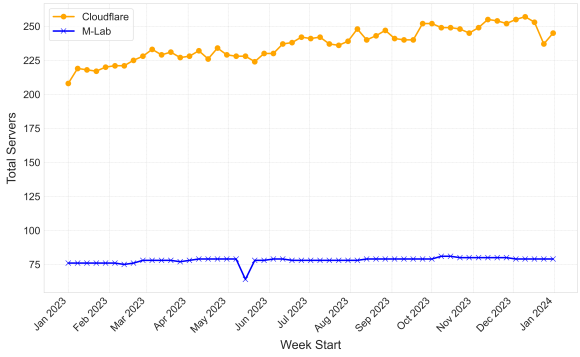
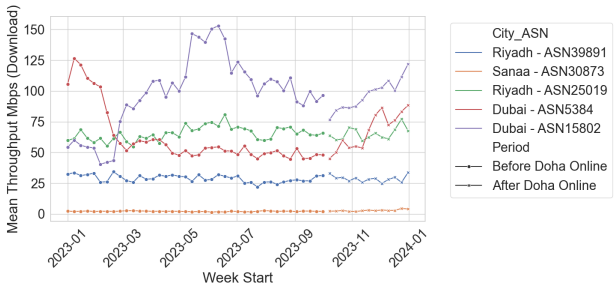


Figure 6: The number of speed test servers used per week.

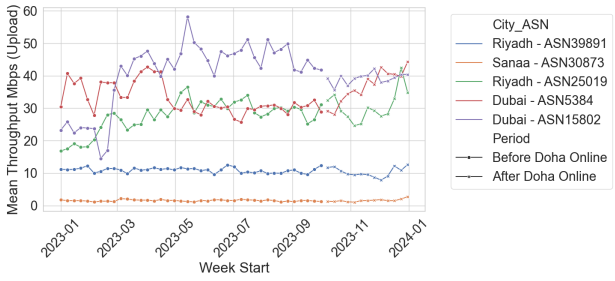
M-Lab allowing for custom server selection, which Cloudflare does not [20].

However, as Figure 6 shows, due to its CDN nature, Cloudflare operates significantly more servers than M-Lab, likely deploying at existing locations or reusing infrastructure. Cloudflare operates over 300 server locations globally throughout the year, approximately 4 times more than M-Lab's deployment. M-Lab uses dedicated servers and relies on third parties for deployment [15]. This fundamental difference affects both geographic coverage and performance characteristics. M-Lab server count remains mostly constant around 75 throughout the year. They add 5 server locations during the year, likely new deployments given their 3-month appearance delay. For Cloudflare, the nature of new servers is less clear due to low test numbers and many servers.

To investigate the impact of adding a new server on regional results, we examine M-Lab data.



(a) Download



(b) Upload

Figure 7: M-Lab download and upload speeds for City-ASN pairs using the Doha server.

We selected one of the additions to M-Lab, a new server located in Doha, Qatar, showing up in the data for the first time on 08.10.2023. We then search for city-ASN pairs that use this Doha server after its appearance, and look at what servers they were using before. This results in Figure 7a, showcasing not a sudden jump due to the new location, but an upwards trend in tested speed. Furthermore, we also inspected the adoption rate of the newly launched server. We see some locations, such as Dubai, adopt the new server exclusively, due to geographical proximity. Other locations continue to use the servers they used previously, but also run tests via the new Doha server. Under 5% of city-ASN pairs proceed to use the new server location exclusively, while a much larger percentage runs tests from other locations as well as the new option.

Key takeaways: The addition of a new test location in a region that previously did not have a geographically close testing facility showcases a consistent upwards trend of tested speeds. Clients usually adopt the newly added server location in addition to the previously used one, with only 5% of geographically very closely located clients using the new location exclusively.

6.5 Latency & loss rate

Figure 8 shows the CDF of the minimum latency of both Cloudflare and M-Lab results across the entirety of 2023. While Cloudflare gives access to an array of latency values, M-Lab only provides the minimum latency value. Therefore, we only compare on that

basis. As expected, due to the higher number of servers, Cloudflare latency is slightly lower than M-Lab's across the board.

We do not compare loss rate between test providers, as M-Lab uses TCP connection lifetime loss rate and Cloudflare uses a separate UDP based loss rate test.

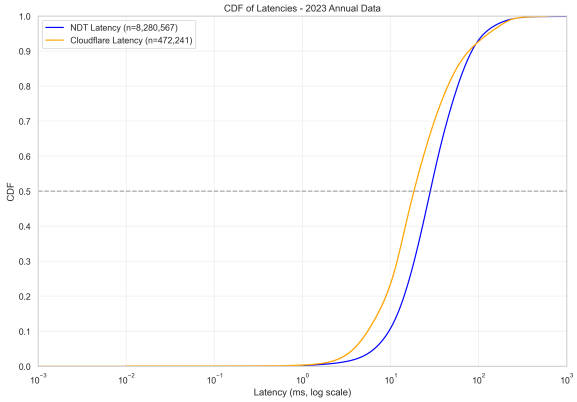


Figure 8: CDF comparison of latency values in 2023.

Key takeaways: Cloudflare latency is slightly lower than M-Lab's due to their higher number of servers. M-Lab only gives access to minimum latency, while Cloudflare gives an array of values, comparisons therefore have to be done on the basis of the common value, in this case minimum latency. Loss rate reported by M-Lab and Cloudflare are not comparable.

6.6 IP version throughput differences

Additionally, we also examined throughput differences per IP version. Previous results shown in this paper make no distinction between IPv4 or IPv6 speed tests. Due to the large amount of BigQuery data consumed, this comparison was only done for the month of January 2023. Figure 9 shows both a scatter plot and CDF plot of this data, aggregated on a weekly basis, without applying the previously mentioned minimum of 3 tests per week, split by IP version. Table 3 from earlier in this paper shows the exact number of city-ASN pairs present for each version. Although the throughput values reported by M-Lab outperform those reported by Cloudflare in both tests (similarly to Figure 5a) the difference is more pronounced in IPv6 tests especially for download speeds. In addition, when considering the median global performance difference, IPv6 performs comparably in upload speed, but a much more pronounced difference can be seen in download speed.

Key takeaways: Speed tests done using IPv6 have a significantly increased throughput performance when compared to tests performed using IPv4, when considering the global scale. This difference is especially notable in download throughput and less pronounced in upload performance.

7 Guidance for Researchers

Based on our analysis, we provide practical guidance for speedtest data usage:

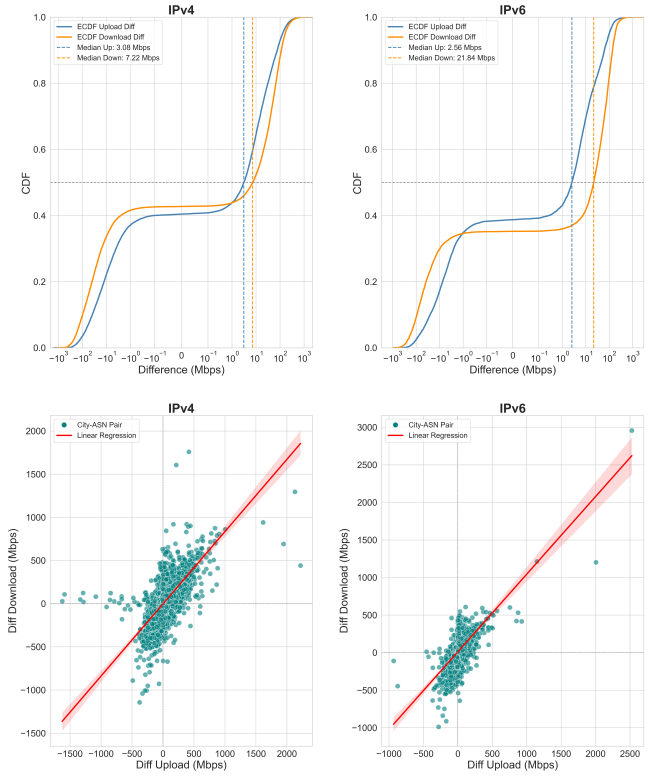


Figure 9: January 2023 CDF and scatter plot, split by IP versions

Use M-Lab for: Long-term trends (since 2009), single-stream TCP analysis, large samples (10:1 ratio), developing regions, consistent methodology.

Use Cloudflare for: Modern web performance, CDN-like infrastructure, edge measurements, IPv6 studies.

Use both for: Methodology validation, bias studies, comprehensive analysis.

Key considerations: Account for 10:1 sample size difference, expect 5-10% systematic bias, apply robust filtering, use longer aggregation periods.

8 Conclusion

We examined throughput measurement differences between M-Lab and Cloudflare in 2023. We found a statistically significant difference in throughput measurements between the platforms, within the 5-10% range for the global average (Section 6.3). This systematic bias reflects fundamental methodological differences rather than random measurement errors. We also found that there are countries where this difference is more pronounced, such as Switzerland and Armenia (Section 6.2), primarily due to asymmetric server deployment strategies. Additionally, we discovered that there is a large difference in quantity of measurements recorded between the two platforms, and that before working on this data, an interested party needs to exercise caution to preprocess it to eliminate malformed and incomplete measurements (Section 5.4). Furthermore, we discovered that there is an observable difference in speed test performance between tests done using IPv4 and IPv6, with IPv6 outperforming IPv4 especially in download throughput.

We also examined what impact the addition of a new testing server location has on speed test results, finding its addition to have a consistent upwards impact on tested speeds for tests done using the new location.

References

[1] Eneko Atxutegi, Fidel Liberal, Eduardo Saiz, and Eva Ibarrola. 2015. Why we still need standardized internet speed measurement mechanisms for end users. In *2015 ITU Kaleidoscope: Trust in the Information Society (K-2015)*. 1–8. doi:10.1109/Kaleidoscope.2015.7383645

[2] Vaibhav Bajpai and Jürgen Schönwälder. 2015. A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *IEEE Communications Surveys & Tutorials* 17, 3 (2015), 1313–1341. doi:10.1109/COMST.2015.2418435

[3] Steve Bauer, David Clark, and William Lehr. 2010. *Understanding Broadband Speed Measurements*. Technical Report. Massachusetts Institute of Technology.

[4] Caio. 2020. Server Selection for the New Speedtest. https://community.cloudflare.com/t/server-selection-for-the-new-speedtest/177498. Accessed: 1 October 2024.

[5] Marshini Chetty, David Haslem, Andrew Baird, Ugochi Ofoha, Bethany Sumner, and Rebecca Grinter. 2011. Why is my internet slow? making network speeds visible. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1889–1898. doi:10.1145/1978942.1979217

[6] Cloudflare. 2025. Test speed · Cloudflare Fundamentals docs. https://developers.cloudflare.com/fundamentals/performance/test-speed/ Accessed: 2024.

[7] Federal Communications Commission. 2023. Measuring Broadband America. https://www.fcc.gov/general/measuring-broadband-america. Accessed: 1 October 2023.

[8] Measurement Lab Community. 2024. Discussion Thread. https://groups.google.com/a/measurementlab.net/g/discuss/c/lV4HK2QRM4A?pli=1. Accessed: 1 October 2024.

[9] Xiaohong Deng. 2018. *Comparing ISP Broadband Performance Using M-Lab Data*. Master's thesis. University of New South Wales. doi:10.26190/unsworks/3647

[10] Nick Feamster and Jason Livingood. 2020. Measuring Internet Speed: Current Challenges and Future Recommendations. *Commun. ACM* 63, 12 (December 2020), 72–80. doi:10.1145/3381039

[11] KNawm. 2022. Speed Test Using Cloudflare Workers. https://github.com/KNawm/speed-cloudflare-cli. Accessed: 1 October 2024.

[12] Measurement Lab. 2023. BigQuery Quickstart Guide. https://www.measurementlab.net/data/docs/bq/quickstart/. Accessed: 1 October 2023.

[13] Measurement Lab. 2023. Measurement Lab. https://www.measurementlab.net/. Accessed: 1 October 2023.

[14] M-Lab. 2015. NDT Test Methodology. https://github.com/ndt-project/ndt/wiki/NDTTestMethodology Accessed: 2024.

[15] M-Lab. 2023. How to Deploy M-Lab Site. https://andrew-m-lab.github.io/publications/how-to-deploy-mlab-site.pdf. Accessed: 1 October 2024.

[16] Kyle MacMillan, Tarun Mangla, James Saxon, Nicole P. Marwell, and Nick Feamster. 2023. A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7). *SIGMETRICS Perform. Eval. Rev.* 51, 1 (June 2023), 41–42. doi:10.1145/3606376.3593522

[17] Measurement Lab. 2023. Most NDT Clients Migrated to NDT7. https://www.measurementlab.net/blog/most-ndt-clients-migrated-to-ndt7/ Accessed: 2025-02-26.

[18] Measurement Lab. 2023. Speed Test by Measurement Lab. https://speed.measurementlab.net/ Accessed: 1 October 2023.

[19] Al Morton. 2013. Improved Internet Speed Tests Can Enhance QoS and QoE. In *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013)*. IEEE, Vienna, Austria, 115–119. doi:10.1109/PQS.2013.6631737

[20] Netbeez. 2023. Measuring Internet Speed via NDT. https://netbeez.net/blog/measuring-internet-speed-via-ndt/. Accessed: 1 October 2024.

[21] Ofcom. 2023. Home Broadband Performance: March 2023. https://www.ofcom.org.uk/__data/assets/pdf_file/0028/257592/home-broadband-report-march-2023.pdf. Accessed: 1 October 2023.

[22] Lai Yi Ohlsen. 2023. Measurement Lab and Cloudflare Announce Release of AIMScoreData. https://www.measurementlab.net/blog/cloudflare-aimscoredata-announcement/. Accessed: 1 October 2023.

[23] Lai Yi Ohlsen. 2023. Updates to Geo Filters. https://www.measurementlab.net/blog/updates-to-geo-filters/. Accessed: 12 March 2025.

[24] Ookla. 2025. Ookla for Good. https://www.ookla.com/ookla-for-good. Accessed: 2025-07-07.

[25] David Tuber. 2023. AIM Database for Internet Quality. https://blog.cloudflare.com/aim-database-for-internet-quality. Accessed: 1 October 2023.

[26] Fred B. Wood, Victor H. Cid, and Elliot R. Siegel. 1998. Evaluating Internet End-to-end Performance: Overview of Test Methodology and Results. *Journal of the American Medical Informatics Association* 5, 6 (11 1998), 528–545. doi:10.1136/jamia.1998.0050528 arXiv:https://academic.oup.com/jamia/article-pdf/5/6/528/2082608/5-6-528.pdf