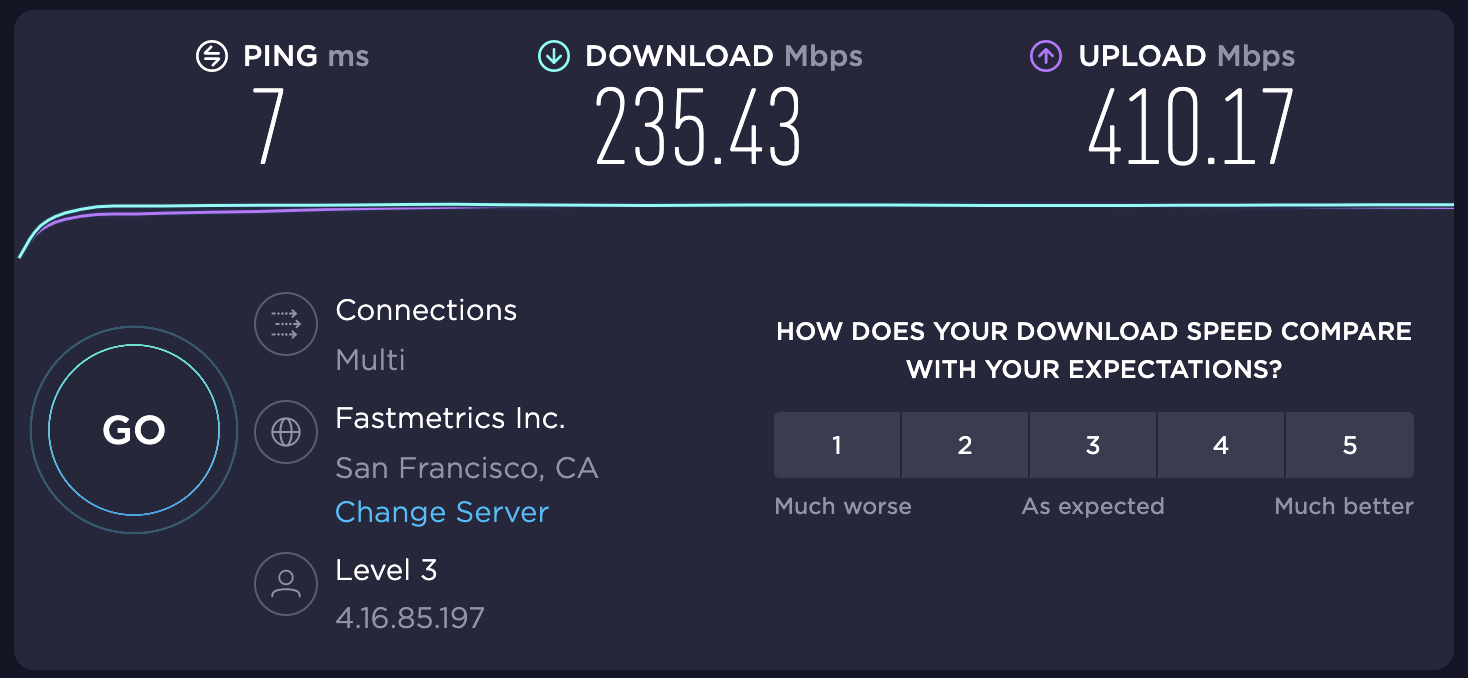
AIM Score

When you fire up speedtest.net and run your speed test, you get a response that (hopefully) looks something like this:



But from a customer perspective, what does this mean? Nowhere on the speedtest site does it say what each of these values is and how they correlate to ISP performance. Even in the wiki, [bandwidth is incorrectly attributed to speed](https://www.speedtest.net/about/knowledge/faq#streaming). This can lead to people mistakenly inferring that bandwidth is the only factor in determining how "good" their Internet connection is.

In truth, there are many factors that actually determine a user's Internet performance. Distance from content, bandwidth, relative shape of the path, packet loss, and origin reachability all factor into Internet quality, as well as things that are often abstracted from customers, like routing security, NXDOMAIN pollution, and bandwidth mismatches causing things like bufferbloat. While it's useful to produce all of these values to customers, the amount of information can be too much for people without domain knowledge to process if not done correctly. Therefore it's useful for us and for customers if the Internet community was able to produce a singular score that took all of these factors into account to produce a score that highlights Internet performance while giving customers one simple number that they can use to compare their experience with others on their network or in the area.

This document introduces the concept of an aggregated Internet measurement, or an AIM score. This score uses a point system to assign value to different characteristics of an end user's Internet performance, and then aggregates them into a single number score out of 100.

### **Calculating AIM Score: What matters?**

The first principles of any analysis of last mile Internet performance comes down to two things: performance and availability. If a provider is not able to reach web sites reliably, it is not fulfilling its basic premise of providing Internet connectivity. If a provider is not performant, this will result in users becoming unsatisfied and may lead to churn. Therefore, any score must weight availability and performance metrics more heavily than other metrics.

For some applications, loss rate and jitter heavily influence availability, so they become an availability factor as well. VoIP services require low jitter and low loss paths, so being able to properly position networks by these pivots is also important.

Performance metrics are comprised of a latency measurement and a throughput measurement.

1. The goal of the latency measurement is to highlight how good a provider is at reaching origin networks quickly. Therefore, we will use min RTT, as that is a good approximation of physical distance from an origin.
2. The goal of the throughput measurement is to show how big the "pipes" are between the content and the user. Larger numbers can be better, but different bandwidth-RTT numbers can have [similar BDP values](https://en.wikipedia.org/wiki/Bandwidth-delay_product) so it's important that we take that into context.

### **What matters but less than availability and performance?**

There are many initiatives on the Internet that help improve quality of life for users that don't directly translate to the above categories, but have definite impact on the user experience. For example, RPKI support, NXDOMAIN pollution protections, and Path MTU discovery are also important in evaluating a network, but don't directly translate to availability or performance on a consistent basis.

### **Score calculation process**

End-user flow:

1. User connects to the site (speedtest, warp, whatever)
2. User performs the following across 5 different origin sites:
   1. ICMP ping (or similar)
   2. Download a small file
   3. RPKI check
   4. check to see if site is preventing NXDOMAIN pollution
   5. Path MTU discovery
3. Data is aggregated to form score
4. Score is presented, along with compares of other users on similar networks in the area, also showing where user falls in percentiles

### **Assigning points to each metric**

This table shows the point thresholds for each of the different metrics that are present as a part of the AIM score.

| **Metric** | **0 points** | **5 points** | **10 points** | **20 points** | **30 points** | **50 points** |
| --- | --- | --- | --- | --- | --- | --- |
| IPv4 reachability | down |  |  |  |  | up |
| IPv6 reachability | down |  | up |  |  |  |
| Loss Rate | > 10% |  | < 10% | < 5% | < 1% |  |
| Jitter | > 20 ms | < 20ms | < 10ms |  |  |  |
| RTT | > 100ms | < 100ms | < 10ms |  |  |  |
| Download Throughput | < 1Mbps | < 10Mbps | < 50Mbps | < 100Mbps | < 1000Mbps |  |
| Upload Throughput | < 1Mbps | < 10Mbps | < 50Mbps | < 100Mbps | < 1000Mbps |  |
| UL-DL percent difference | > 50% | < 50% |  |  |  |  |
| RPKI support | No |  | Yes |  |  |  |
| NXDOMAIN pollution | No | Yes |  |  |  |  |
| Path MTU Discovery | No | Yes |  |  |  |  |

For example, if a user had the following scenario

* City - Boston
* ASN - 701
* Origin - NYT
* RTT - 5
* DL - 10Mbps
* UL - 10
* Loss Rate - 2%
* Jitter - 21ms
* IPv4 - 1
* IPv6 - 1
* RPKI - 1
* NXDOMAIN - 1
* PMTUD - 1

Their total score would be 135 points. We can then normalize that to some number out of 100, which is then presented as a comparison with other Boston users, as well as total percentile of internet users in the world.

The maximum total score any network can have is 185, which means the normalized score for this user is 70.

Another example:

* City - Seattle
* ASN - 11404
* Origin - NYT
* RTT - 1ms
* DL - 311Mbps
* UL - 300Mbps
* Loss Rate - 0%
* Jitter - 0ms
* IPv4 - 1
* IPv6 - 1
* RPKI - 1
* NXDOMAIN - 1
* PMTUD - 1

This score is 185 points, giving this user a normalized score of 100. This user would be described as having perfect Internet to this origin.

### **Second order metrics**

Once the base metric values are calculated, we can use those to form second order metrics which can tell us a bit more about the ISP's behavior across multiple endpoints

* Traffic Shaping metric: This metric calculates the difference in throughputs across different origins. In a perfect network, each origin should have the same throughput, but some ISPs will throttle different origins to have slower throughput, which will be reflected in the difference in download bandwidth across the origins
* UL - DL difference: This metric calculates the difference in upload and download throughput. In a perfect network, the difference between these two numbers should be negligible (i.e. less than 50% difference). Some networks may throttle upload speeds, which will provide a poor customer experience

These metrics are collected across each of the origins and will combine to form a singular score. This is done by taking the average and variance across all the different metrics and evaluating them as one aggregate score. This score is then presented to customers along with percentile values and comparisons to other users in the region and other users on that network.

##### **What if my Wi-Fi is making things worse?**

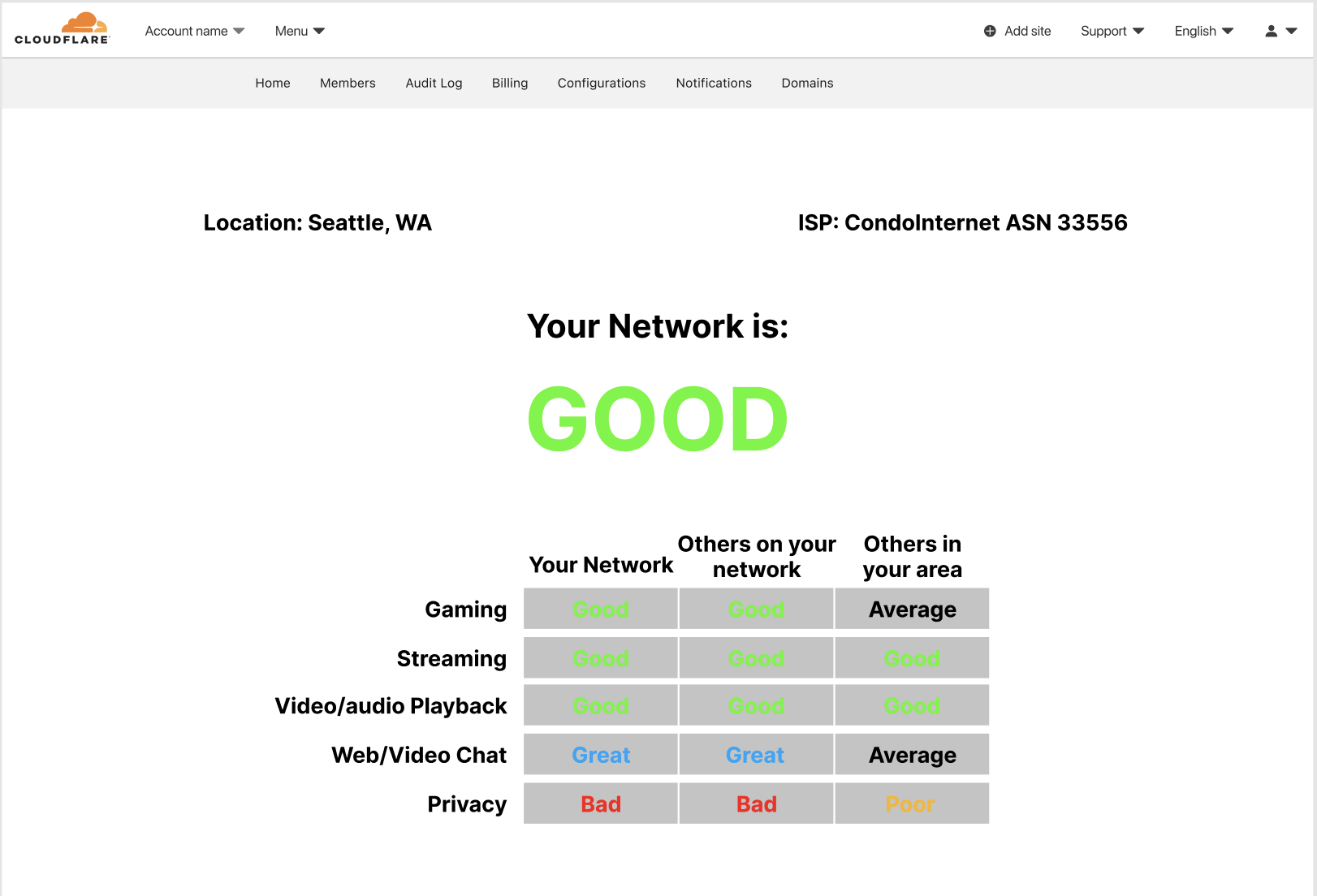
As noted in external partner conversations, specifically from [ISPs](https://www.iab.org/wp-content/IAB-uploads/2021/09/xfinity-wifi-ietf-iab-v2-1.pdf), end-user Wi-Fi placement greatly impacts scores, and could actively lower an ISP's metrics if they count too many scores which are impacted by bad Wi-Fi. Therefore, to help mitigate this, if a user is reporting a score for an ISP that is markedly lower than other scores from that provider, we could display a notification that the speed test could be due to poor performance on the Wi-Fi router and to try again with either a direct connection, or better placement in relation to the access point.

### **Third order metrics: What is my network good for/how good am I compared to others like me?**

One important drawback of existing numerical scores is that they don't necessarily do a good job of figuring out what applications perform better on a particular connection. To quote one user: "One of the first things I did during the pandemic was attempt to upgrade my Internet bandwidth, but my videoconferencing still sees drops." This is indicative of bandwidth not being a good measurement of network quality or user experience. To that end, any network score should allow customers to pivot on different common use cases that would perform better or worse on their connection. The following use cases could be considered as things users would want to optimize for:

* Gaming
* Streaming
* Video/audio playback
* Web chat/RTC
* Privacy

Each of these scenarios would have different network metrics that would need to be at certain thresholds for the network to be "good" for that particular application. That would then allow users to compare their service to others using the same provider, or even users in the same region:



### **Partner requirements**

In order to properly put together a network score that everyone will use, we need multiple partners to sign on and start adopting the AIM score. Key players to reach out to:

* Apple
* Microsoft
* Google
* Google uses an independent speed test called M-Lab
* Ookla (speedtest.net owners)
* Netflix
* Zoom

### **External sources**

IAB keynote speakers broadly [suggest](https://www.iab.org/wp-content/IAB-uploads/2021/09/Lower-layer-performance-is-not-indicative-of-upper-layer-success-20210906-00-1.pdf) that the existing network performance measurements are outdated and need to be re-evaluated: <https://www.iab.org/activities/workshops/network-quality/>

### **Calculate your own ISP score**

Please help contribute to this effort by r[unning speed test and ping tests on your own home network (or mobile)](https://github.com/dtuber/aimscoretest) and reporting your numbers on [this sheet](https://docs.google.com/spreadsheets/d/10OHrU1zD4C4wCtnP3UJsdbdsTXzWkxFGcoQJHPpH28M/edit?usp=sharing). Also in additional notes please add whether or not you subjectively feel like you have good internet (to help us understand whether or not the score is reflective of reality).