Access Network Quality as Fitness for Purpose

Kenjiro Cho (IIJ/WIDE Project)

In this position paper, I discuss issues in using speedtests for network quality, and argue for the use of fit-for-purpose criteria for access network quality measurement.

1. Why are speedtests so popular?

Many of us love speedtests, probably for a number of reasons:
- It is easy to measure at a speedtest site.
- Results are easy to understand (e.g., the bigger the better!, easy to compare).
- Throughput is, technically, the most important metric for network performance.

Therefore, speedtest is popular not only among consumers but also among engineers, media, policy makers, and ISPs. Although speedtest is a nice tool for casual throughput testing, these perceived benefits do not necessarily hold, especially when used for quality measurement.

It seems easy to measure at a speedtest site, but it is hard to measure correctly since results could be influenced by various factors such as home WiFi, CPU load and cross traffic. In addition, measurements fluctuate with non-trivial noise that is hard to filter out with traditional methods.

Results seems easy to understand, but could be extremely difficult to interpret or compare. One needs to know the time, place and environment of measurements, and also needs to understand statistical multiplexing and the nature of fluctuations. Results are likely to vary among different speedtest sites employing slightly different measurement methods, as well as when the bottleneck is not the access link, or service differentiation is enabled; various types of service differentiation will soon be used for popular content and services.

Moreover, speedtest is a stress test in public space which generally requires prudent planning and execution. It could affect other services, even though the impact of speedtest should be similar to downloading large content. Still, it inevitably increases the queueing delay at the bottleneck. Similarly, measurements could be affected by others, and a user cannot get expected results when others sharing the bottleneck also try the measurement.

In addition to being intrusive, speedtest is energy-wasting and not environment-friendly, if it matters. The research community has proposed alternative methods for measuring available bandwidth without actually filling the pipe since 1990s (e.g., packet-pair[1], pathchar[2], pathrate[3]). Unfortunately, they are not as popular as speedtests, and not widely employed even by research efforts.

Throughput is technically the most important metric for network performance, and it is certainly essential to accurately measure throughput for the engineering or operational purposes. Performance however does not exactly mean quality; quality is in general an overall satisfaction index combining multiple performance metrics and other factors.

Contrary to the fact that speedtests are popular and throughput is the only metric well recognized among network performance metrics, it seems not so suitable for representing the quality of access networks.

2. Psychology of speedtests: performance over stability

When a speedtest is used to show the best performance a user can get, it is straightforward; a user can simply pick the best result among multiple trials. When a speedtest is used, however, to show poor performance by a user who is frustrated with poor service quality, it is not as simple as that. A broadband service is a best-effort service shared by other users so that the performance is expected to fluctuate. Thus, a single measurement result cannot prove poor quality, and speedtest sites do not provide a means beyond a single measurement. In short, a speedtest is a nice tool for happy users but not for frustrated users. To improve the quality of broadband services as a whole, we need to identify and quantify underserved users and fix their problems.

Nevertheless, speedtests introduce a psychological bias towards high throughput; people tend to focus on good scores rather than poor scores when looking at numerical results. For example, there would be no practical difference for a broadband user between 600Mbps and 800Mbps, but the latter looks much better. Consequently, speedtests induce a bias among users and ISPs towards raising top-level performance rather than improving poor performance, which could guide the market towards an undesirable direction.

Another issue often addressed is a discrepancy between an advertised access-link capacity and speedtest results. When a user upgrades a broadband service from 100Mbps to 1Gbps, they expects speedtest scores to jump up by a factor of 10, even when the service fee increases by only 10%.

Increasing the link capacity brings two benefits: one is to improve the throughput performance, and the other is to improve the stability by over-provisioning. The former is what users have in mind, but the latter is operationally more important as unused capacity absorbs fluctuations in traffic, and directly contributes to user experiences with video-conferencing, video-streaming and other applications. Unfortunately, we lack an easy-to-understand metric for network stability.

Another problem lies in advertisement of access network services. To differentiate a service in advertisement, many ISPs and carriers emphasize access capacity and service fee. In other words, ISPs and carriers do not have other factors to advertise and, as a result, users think network capacity to be a single criterion. This situation is to some extent similar to automotive industries before the 1973 oil crisis; quality cars were believed to be large, heavy, and powerful so that these specs were featured in advertisements. It is time for us to change the current mindset about network service quality.

3. Quality as fitness for purpose

The fit-for-purpose quality is a concept within Quality Assurance emphasizing to meet the needs of customers. Although there are diverse user needs for broadband services, we can define standard quality criteria that would satisfy the vast majority of users. In a household, for example, it is not uncommon these days to have video-conferencing for each parent and video-streaming for each kid at a time. Assuming 2 parents and 2 kids, it roughly translates to 10Mbps download, 5Mbps upload, with other requirements on delay and jitter. These values would satisfy most households (e.g., 95%), except serious gamers, traders and other minority users with specific requirements so that it would work as standard fit-for-purpose quality criteria. The criteria should be revisited and updated every few years.

A test result for the criteria is binary, whether it satisfies the criteria or not. The overall quality score can be expressed by the availability, that is the success ratio of periodic measurements. For other network services, availability (e.g., 99.9%) is commonly used for Service Level Agreement (SLA) or Service Level Objective (SLO). Thus, it would be consistent to express the fit-for-purpose quality by availability. Then, when a broadband service has 99% availability for the given fit-for-purpose quality, a customer can expect the service to satisfy the standard quality for 99% of time, but may not for 1% of time (14.4 minutes per day) on average.

The above example standard quality can be measured by a tool that emulates video-conferencing and video-streaming traffic patterns, and requires only limited bandwidth. The details of the tool design is out of scope of this position paper, but it would not be too difficult to develop such a tool. To measure the availability, it should obtain enough samples over different times of day and days in week. The considerations for time and population sampling are similar to coordinated throughput data collection and, in fact, it is possible to show fit-for-purpose quality from throughput results. Still, the required accuracy is very different for throughput performance metric and for fit-for-purpose criteria (e.g., accurately measuring 1Gbps vs. 10Mbps), and the variability in fit-for-purpose measurements is much smaller. So, a data collection system for fit-for-purpose criteria can be much simpler with much lower network load and much lower system cost.

5. Summary

Although speedtest is very popular, it may not be the right tool to improve the quality of access network services. Speedtest is not suitable to quantify poor network services, and could be misleading if used for quality measurement. It is time to change the current mindset about network service quality, and look into the use of fit-for-purpose criteria.

References

[1] S. Keshav. A Control-Theoretic Approach to Flow Control. ACM SIGCOMM '91, September 1991.

[2] Van Jacobson. Pathchar: A Tool to Infer Characteristics of Internet Paths. Presented at the Mathematical Sciences Research Institute(MSR1), available from ftp://ftp.ee.lbl.gov/pathchar/, April 1997.

[3] C. Dovrolis, P. Ramanathan and D. Moore. Packet-dispersion techniques and a capacity-estimation methodology. IEEE/ACM Transactions on Networking, vol. 12, no. 6, pp. 963-977, December. 2004, doi: 10.1109/TNET.2004.838606.