A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7)

Kyle MacMillan
University of Chicago
Chicago, USA
macmillan@uchicago.edu

Tarun Mangla
University of Chicago
Chicago, USA
tmangla@uchicago.edu

James Saxon
University of Chicago
Chicago, USA
jsaxon@uchicago.edu

Nicole P. Marwell
University of Chicago
Chicago, USA
nmarwell@uchicago.edu

Nick Feamster
University of Chicago
Chicago, USA
feamster@uchicago.edu

ABSTRACT

Consumers, regulators, and ISPs all use client-based "speed tests" to measure network performance, both in single-user settings and in aggregate. Two prevalent speed tests, Ookla's Speedtest and Measurement Lab's Network Diagnostic Test (NDT), are often used for similar purposes, despite having significant differences in both the test design and implementation, and in the infrastructure used to perform measurements. In this paper, we present the first-ever comparative evaluation of Ookla and NDT7 (the latest version of NDT), both in controlled and wide-area settings. Our goal is to characterize when and to what extent these two speed tests yield different results, as well as the factors that contribute to the differences. To study the effects of the test design, we conduct a series of controlled, in-lab experiments under a comprehensive set of network conditions and usage modes (e.g., TCP congestion control, native vs. browser client). Our results show that Ookla and NDT7 report similar speeds under most in-lab conditions, with the exception of networks that experience high latency, where Ookla consistently reports higher throughput. To characterize the behavior of these tools in wide-area deployment, we collect more than 80,000 pairs of Ookla and NDT7 measurements across nine months and 126 households, with a range of ISPs and speed tiers. This first-of-its-kind paired-test analysis reveals many previously unknown systemic issues, including high variability in NDT7 test results and systematically under-performing servers in the Ookla network.

CCS CONCEPTS

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

KEYWORDS

Speed Test, Ookla, Network Diagnostic Tool, Broadband, Internet Speed, Measurement Lab

SIGMETRICS '23 Abstracts, June 19–23, 2023, Orlando, FL, USA.

ACM ISBN 979-8-4007-0074-3/23/06.
https://doi.org/10.1145/3579448

ACM Reference Format:
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). In *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '23 Abstracts), June 19–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3579448

1 INTRODUCTION

Network throughput—colloquially referred to as "speed"—is among the most well-established and widely used network performance metrics. Indeed, "speed" is used as the basis for a wide range of purposes, from network troubleshooting and diagnosis, to policy advocacy [3, 10, 16, 18] (e.g., on issues related to digital equity), to regulation and litigation [4] (e.g., on issues related to ISP advertised speed). Given the extent to which stakeholders, from consumers to regulators to ISPs, all rely on "speed", it is in some sense surprising that there is no consensus on the way to measure it. Absent any standard, many speed tests, varying in both design and implementation, are used interchangeably.

Over the past decade, Ookla's Speedtest [13] ("Ookla") and Measurement Lab's Network Diagnostic Tool ("NDT") [7] have been widely used by both consumers and policymakers: Ookla and NDT report a daily average of over 10 million [11] and 6 million tests [6], respectively. As a result, the compiled datasets from these two tests, amounting to billions of speed tests [2, 11], have become universal resources for analyzing broadband Internet performance [4, 10, 16, 18]. Unfortunately, these datasets have also been used out of context, without a clear understanding of the caveats and limitations of these tools under different circumstances and environments [12].

The stakes—and, therefore, the costs—of misuse have also never been higher. In the United States, Congress has committed $43.5 billion to Internet infrastructure, including to last-mile performance and availability improvements [1]. In response, state and local officials across the country are currently urging consumers to participate in speed test crowd-sourcing initiatives to help establish which areas meet the federal funding criteria [14].

To their credit, the organizations who have developed these speed test tools have tried to prevent misappropriation of the data by issuing guidance about how the tools and public data should and should not be used. M-Lab has gone as far as to say that "*M-Lab's NDT and Ookla's SpeedTest measure fundamentally different things*" [8].

While this statement is certainly true, there has been no study to date about how these differences in tool design can (and do) yield different results in practice, under different operating conditions. Acknowledging that Ookla and NDT7 are different is, in some sense, besides the point. Although each tool may have been designed with a specific purpose in mind, that does not mean it can not fulfill—or be appropriated for—other purposes. Such has been the case with NDT7, which has been used as a tool to measure access ISP throughput, even though its stated design is to test the throughput of *a single TCP connection*. In light of the significant attention to both of these tests, it is imperative to develop a rigorous, quantitative, and specific understanding of the circumstances under which each tool can accurately measure last-mile speed—and, hence, the context for interpreting each dataset.

To this end, we conduct the first-of-its-kind systematic, comparative study of NDT7 (the latest version of NDT) and Ookla[1]. We begin with a set of in-lab experiments that allow us to directly compare the tools under controlled network conditions where "ground truth" is known. Next, we conduct more than 80,000 paired wide-area network tests, whereby the two tests are run back-to-back, from 126 home broadband access networks across more than 30 neighborhoods in one of the largest cities in the United States for nearly a year. In-lab, we use controlled experiments to characterize how NDT7 and Ookla behave under a wide range of network conditions—specifically, varying throughput, latency, packet loss, and cross-traffic. We also study how different transport congestion control algorithms and client types (i.e., browser vs. native client) may affect the measurements that each tool reports. Second, we compare the behavior of these two tools using data from our wide-area network deployment encompassing 10 different ISPs. A unique and important methodological aspect of our study is the use of *paired speed tests*, where we run Ookla and NDT7 in succession. To our knowledge, this is the first comparative analysis of Ookla and NDT7 in deployment over a significant number of networks for an extended period of time.

2 MAIN FINDINGS

Our main findings are summarized below:

- The NDT7 client can send at about 95% of a high-capacity link (up to 2 Gbps) using only a single TCP connection. This finding updates past work that reported a different finding, that a single TCP connection can not achieve a throughput approaching full capacity [5].
- The NDT7 client under-reports throughput at higher latencies, in comparison to Ookla: The Ookla client reports speeds up to 12% higher than NDT7 at 200 ms round-trip latency, and up to 56% higher at 500 ms latency.
- Across all households in the wide-area deployment, the median fraction of paired tests for which Ookla reports a speed that is 0–5% higher than NDT7 is 73.8%. The fraction of paired tests for which Ookla reports a speed that is 5–25% higher is 13.4%.
- For Ookla, the choice of test server can significantly affect the reported speed. Tests using certain Ookla servers systematically report speeds 10% lower than other servers.
- NDT7 tests are more likely to under-report during peak hours. 43.4% of households observed a statistically significant decrease in NDT7-reported download speed tests during peak hours, whereas only 18.9% of these households saw the same for Ookla.

[1] We focus on Ookla and NDT7 because of their popularity with consumers and policy makers, but the method in this paper also applies to other tools.

3 CONCLUSION

This work provides an in-depth comparison of Ookla and NDT7, focusing on both test design and infrastructure [9]. Our results and suggestions help users understand why the reported speed from Ookla and NDT7 may differ, as well as guide policymakers towards more accurate and appropriate use of Ookla and NDT7 data. We have released all collected data, as well as all of the measurement code we used to conduct the study [15, 17].

Although this is the *first* work to perform a controlled and extensive comparison of Ookla and NDT7, we neither hope nor expect that it will be the last, as many important technical and policy questions still remain. In the paper, we the implications of our findings for the future of speed test tools and data analysis and outline multiple avenues for future work. We view this research as the beginning of a discussion on how to use collective speed test data to shed more light on the state of broadband Internet access networks around the United States, and the world.

REFERENCES

[1] Josh Boak. 2022. Biden Administration to Release $45B for Nationwide Internet. https://apnews.com/article/technology-broadband-internet-643949e239296f4f5154509d4c40de73
[2] David D. Clark and Sara Wedeman. 2021. *Measurement, Meaning and Purpose: Exploring the M-Lab NDT Dataset.* SSRN Scholarly Paper. Rochester, NY.
[3] Federal Communications Commission. 2021. Measuring Broadband America Program. https://www.fcc.gov/general/measuring-broadband-america
[4] Federal Communications Commission. 2022. FTC Takes Action Against Frontier for Lying about Internet Speeds and Ripping Off Customers Who Paid High-Speed Prices for Slow Service. https://www.ftc.gov/news-events/news/press-releases/2022/05/ftc-takes-action-against-frontier-lying-about-internet-speeds-ripping-customers-who-paid-high-speed
[5] Nick Feamster and Jason Livingood. 2020. Measuring internet speed: current challenges and future recommendations. *Commun. ACM* (2020).
[6] Measurement Lab. 2020. BigQuery QuickStart. https://www.measurementlab.net/data/docs/bq/quickstart/
[7] Measurement Lab. 2022. Speed test by Measurement Lab. https://speed.measurementlab.net/
[8] Measurement Lab Mailing List. 2017. Measurement Labs (NDT) Speedtest is incorrect? https://groups.google.com/a/measurementlab.net/g/discuss/c/vOTs3rcbp38
[9] 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). *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 7, 1 (2023), 1–26.
[10] New York State Office of the Attorney General. 2020. New York Internet Health Test. https://ag.ny.gov/SpeedTest
[11] Ookla. 2020. About Ookla SpeedTests. https://www.speedtest.net/about
[12] Ookla. 2020. Case study: Ookla and NDT data comparison for upstate New York. https://www.ookla.com/articles/better-funding-decisions-accurate-broadband-network-data
[13] Ookla. 2022. Speedtest by Ookla - The Global Broadband Speed Test. https://www.speedtest.net/
[14] Idaho County Free Press. 2022. Push on internet speed tests to increase chance at 1B USD in funding. https://archive.ph/RWdY8
[15] Ranya Sharma, Tarun Mangla, James Saxon, Marc Richardson, Nick Feamster, and Nicole P Marwell. 2022. Benchmarks or Equity? A New Approach to Measuring Internet Performance. *A New Approach to Measuring Internet Performance (August 3, 2022)* (2022).
[16] Battle For the Net. 2022. Internet Health Test based on Measurement Lab NDT. https://www.battleforthenet.com/internethealthtest/
[17] Internet Equity Initiative UChicago. 2022. Speed test data collected from Netrics. https://github.com/chicago-cdac/netrics-data

[18] Pennsylvania State University and Measurement Lab. 2019. Broadband Availability and Access in Rural Pennsylvania. https://www.rural.pa.gov/publications/broadband.cfm