A First Look at the Spatial and Temporal Variability of Internet Performance Data in Hyperlocal Geographies

Taveesh Sharma, Jonatas Marques, Nicole Marwell, Nick Feamster
University of Chicago

Abstract

Measuring Internet access network performance has been a persistent challenge for researchers and policymakers alike. Unfortunately, existing "speed test" datasets typically lack comprehensive data across both space and time. Specifically, our past work has highlighted that tools like Ookla's Speed Test and Measurement Lab's NDT rely heavily on convenience samples (user-initiated tests from self-selected participants), resulting in a sample that may not generalize across either time or geography. Our ongoing research seeks to address these issues by developing innovative sampling methods and statistical models to provide a more holistic view of Internet performance. Initial findings, focusing on end-to-end latency across hyper-local regions within a single large city in the United States (Chicago, Illinois), reveal that spatial proximity often does not correlate with simultaneous performance anomalies. These insights underscore the need for advanced methods to generalize Internet performance data across time and space. Improved methods can ultimately enable a better understanding of the effects of infrastructure investments on the evolution of Internet performance.

1 Introduction

Measuring the performance of Internet access networks has been a longstanding challenge among Internet measurement researchers, Internet service providers, and policy advocates. Over the past decade, research has made significant advances in developing techniques to measure the performance of a single broadband Internet access link, through the development, evaluation, and comparison of Internet "speed test" tools, in both controlled and wide-area settings. As large datasets on Internet performance from these tools have become more widely available, researchers have begun to use them for a much broader set of purposes. For example, beyond the original goal of measuring the speed of an individual access link, many people are now attempting to ask questions about how broadband Internet access networks perform both over time and across a geography.

Data from two widely used speed tests that have often been used in an attempt to answer these questions, Ookla's Speed Test and Measurement Lab's Network Diagnostic Tool (NDT), both lack systematic longitudinal measures of performance, because the tests are only performed on demand (e.g., when a user is experiencing a performance problem). Thus, temporal assessments of performance are based on ad hoc decisions of users to run speed tests at particular moments in time. These datasets also lack systematic measures of performance across a geography, over-representing certain geographies (as our research shows, often more affluent regions) while under-representing others. Such geographic bias is often incredibly pronounced: For example, many regions in large metropolitan areas in the United States contain no speed test measurements at all. And, while the Federal Communications Commission's (FCC) Measuring Broadband America program captures longitudinal samples of some Internet service provider (ISP) performance, its spatial sampling is even more sparse, sometimes containing only a single measurement vantage point in an entire metropolitan area.

This lack of temporal and spatial granularity in existing datasets brings us to the goal of our ongoing research. Ultimately, we aim to develop new sampling approaches and statistical models that will provide a more comprehensive, representative view of Internet performance over time and across a geography. Our prior and ongoing research over more than a decade, has demonstrated that measurement devices deployed in real-world settings can produce reliable "point data" on Internet performance from individual vantage points. As researchers and the public in general seek to understand Internet performance across regions, however, we face two distinct challenges: (1) understanding how to use performance metrics at a single vantage point to make generalizations about the performance of that access network over time; (2) understanding how to use spatially correlated Internet measurements, from a collection of vantage points in broadband Internet access networks, to make generalizations about Internet performance within a geographic region, and across geographic regions.

Although this general task is extremely challenging, the dataset we have gathered across more than 30 neighborhoods in Chicago over the past several years offers an opportunity to begin asking these questions. In this paper, we begin asking these questions from the perspective of one Internet performance metric: end-to-end latency, both to the last mile and to other destinations on the Internet. Our research makes two observations:

1. Network vantage points that share a persistent increase in end-to-end latency may not always be geographically close to one another.
2. Network vantage points that are geographically close to one another may not always experience the same increase in end-to-end latency.

In other words, spatial proximity—even very close spatial proximity—is often a poor predictor for performance anomalies that may coincide in time. These initial findings, albeit for a simple metric (latency), suggest that many existing samples of Internet performance may not generalize well, either across space, or time.

Ultimately, the research and policy communities should work together to develop new approaches for temporal and spatial sampling, as well as new statistical models that can be used to make generalizations about existing datasets concerning Internet performance. Ultimately, more advanced methods will help us answer questions about how the current and planned investments in Internet infrastructure affect Internet performance in certain regions, and how performance in a particular region evolves over time.

2 Background

Many metrics can be used to evaluate Internet access performance. One of the most common is "speed" (or throughput). Usually estimated by speed test tools, this metric expresses the rate at which data can be successfully transferred between two endpoints. Another important but sometimes overlooked metric is latency—the time it takes for any single bit of data to travel between two network endpoints. In networking, the latency observed for a packet is a result of many delays such as those due to physical propagation as well as packet processing and queuing [8]. The continued rise of near-real-time applications (e.g., video streaming and conferencing, remote surgery, and algorithmic trading) imposes increasingly strict requirements on Internet access latency. Past studies [2] show that even small increases in latency may lead to significant loss in throughput.

In this paper, we focus on round-trip time (RTT) "latency" measurements. We focus on this metric due to the low-cost nature of the measurement, which has allowed us to collect many samples across time and space within a single hyper-local geography. Speed tests, in contrast, consume large amounts of data and thus cannot be performed as often, and are also typically restricted to targeting a limited set of test servers. Many emerging real-time applications, from gaming to videoconferencing, also rely heavily on latency, and many ISPs (e.g., Comcast) are focusing heavily on reducing latency across their access networks as a result. Tools measuring RTT, such as `ping`, send an Internet Control Message Protocol (ICMP) probe to a remote node, which responds with a response probe. The initiating host records both the timestamp when the ping probe was sent and when the response probe was received. The time difference between both timestamps estimates the RTT between the two end-points.

Round-trip latency depends on many factors, such as those due to physical propagation of data, packet queuing at intermediate nodes, and probe processing. Changes in the measured round-trip latency can thus be attributed to changes in these delays. However, there are many confounding factors that can influence the perceived changes to each delay. These factors are related to the path taken to reach the test target, the presence and rate of background traffic sharing the path, and network operation policies (e.g., traffic prioritization), to cite a few. We note that these factors vary both across space and time. Across space because Internet infrastructure is heterogeneous, where some areas may be better provisioned than others [4]. Across time because Internet utilization varies depending on the time of day, week, month, year, and so on [2]. Further, these factors have an impact not only on latency but also on speed tests.

Detecting and measuring the effect of these factors is both important to understand the level of impact that each has on performance and to help explain the obtained results for a particular vantage point in time and space. Understanding these factors and how individual portions of the Internet infrastructure ultimately affect latency is crucial to guide the continued investment in Internet infrastructure. Consequently, we argue that in order to be able to make generalizations on performance over time and across space, it is imperative to perform auxiliary fine-grained measurements that help us better understand the *variability* of these metrics across geography (e.g., across individual vantage points in a city) and time (e.g., from hour to hour or day to day). Measurements concerning latency, including both end-to-end and hop-by-hop, can shed light on these important questions. In the remainder of this paper, we provide initial evidence where latency measurement can lead to meaningful conclusions in this regard; we focus in particular on the challenges in measuring latency across space and time, noting in particular that temporal variability in latency can correlate poorly with geographical proximity.

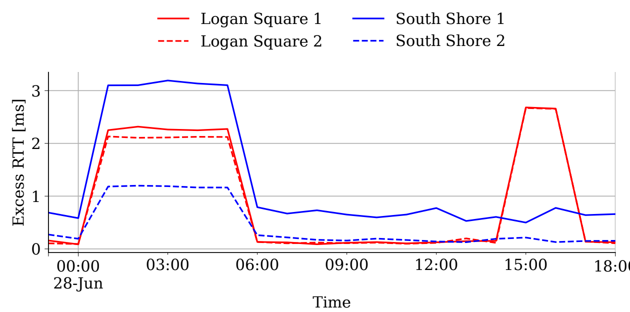
3 Dataset

We used the *Netrics* platform to conduct measurements for this study [6]. The measurements are executed on Raspberry Pis (RPis) connected directly to the home router for residential broadband connections. Netrics' latency measurements are conducted against multiple targets, including fixed distant measurement lab (M-Lab) [5] servers, the ISP's last mile IP address, and popular servers that host services like news, social media, shopping, web search, etc. Measurements to the ISP's last mile involve two steps. First, we conduct a traceroute measurement to a known destination and record the first public IP address along the network path. Then, we conduct a latency measurement to this IP address. For the purpose of analysis in this work, we only consider measurements to fixed IP addresses like M-Lab servers and the ISP's last mile. This is because popular services are often hidden behind load balancers and proxies which leads to unreliable differences in Internet performance. Each latency measurement to a given IP address consisted of 10 ICMP probes, sent every 0.25 seconds with an expiration timeout of 5 seconds each. The average round trip-time over the complete measurement was recorded along with the device demographics, the target server and the timestamp of the measurement. We collected one latency measurement from each device every 5 minutes over the course of the study.

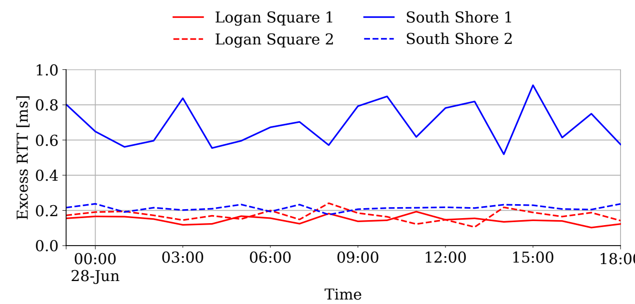
Netrics has been collecting these RTT measurements over more than 2 years till the present day. All of our measurement data is publically available at https://github.com/internet-equity/netrics-data. For our initial analysis, we focus on a sample time period from May 1 to June 30, 2022, when Netrics was deployed on 88 devices, the maximum number of devices we observed at any time. Within this data, we mainly focus on RPis located in two Chicago neighborhoods—Logan Square and South Shore—as they contained the majority of RPis. The choice of these neighborhoods is motivated by a history of socio-economic differences. Both these neighborhoods also do not share a geographical boundary, so we expect them to host distinct regional and local points of presence of the ISP infrastructure. We use data from the remaining neighborhoods only to reaffirm our findings.

4 Preliminary Results

In this section, we describe our preliminary findings of variability of latency measurements across space and time. Our findings from this initial study are preliminary and, to this point, largely anecdotal, but they suggest that the community has a lot of work to do in this area. In particular, when latency exhibits temporal variability, it *does not necessarily mean that endpoints in the same neighborhood are experiencing the same latency variations.* We also find that latency variations can be highly localized, even within the same neighborhood. Although our findings in this paper are largely anecdotal, future work will entail a more systematic study of how latency anomalies within a given geography may diverge.



(a) Latency to Atlanta for Logan Square and South Shore devices



(b) Latency to the ISP's Last-Mile for Logan Square and South Shore devices

Figure 1: Excess RTT time-series slices for Atlanta and the ISP's last-mile respectively. All 4 devices subscribed to the same ISP during the analysis. Each color indicates a different neighborhood.

Result 1: Devices sharing a persistent end-to-end latency spike may not always be located in the same neighborhood. Figure 1a shows a slice of the time series for round-trip latency to an M-Lab server located in Atlanta. There are four devices shown in this figure, with the red devices located in Logan Square, and the blue devices in South Shore respectively. All devices are subscribed to the same ISP. *Excess RTT*, defined as the difference of RTT from the minimum RTT across a time window, is plotted against time. This is done to ensure better visibility into the spikes and ac-

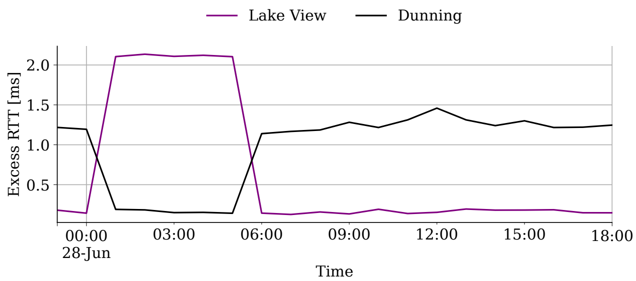
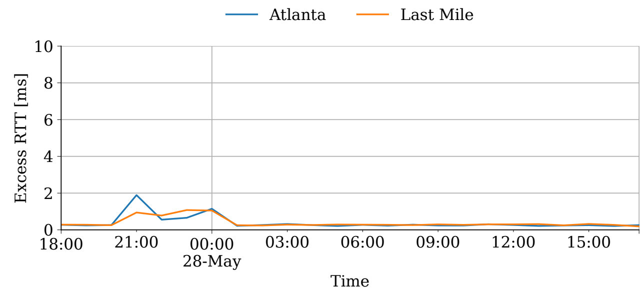
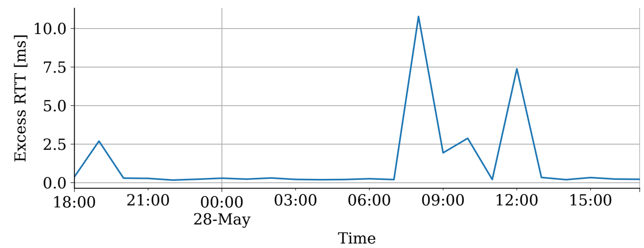


Figure 2: Latency to Atlanta for Lake View and Dunning devices



(a) Latency time-series slice for Atlanta and the ISP's last-mile respectively for `South Shore 2`.



(b) Last-Mile time-series slice for `South Shore 3` – a device located 53 meters from `South Shore 2`.

Figure 3: Plots showing that last-mile latency spikes do not co-occur for devices located close to each other.

curate comparisons with the last-mile latency. We observe two different durations in which the latency to Atlanta rises above the baseline for more than a single measurement. The first change occurred between 12 am and 6 am on June 28, 2022. This change is observed across all four devices, suggesting that it occurred at a link that is shared along the network path to Atlanta for both Logan Square and South Shore. The second change occurred between 3 pm and 6 pm on the same day but only for two devices, both located in Logan Square. This indicates that it was due to the occurrence of network congestion at a link close to or within Logan Square. To evaluate whether either of the changes occurred near the ISP's last-mile access router, we look at the last-mile latencies. Figure 1b shows that the last-mile latency was relatively stable, suggesting that none of these changes originated at the ISP's last-mile.

We also verified if both these changes were only a coincidence. To do this, we looked at the latency to Atlanta for two other neighborhoods – Lake View and Dunning, that do not share a geographic boundary with either Logan Square or South Shore. Figure 2 shows that the first spike was also observed in a device located in Lake View. Contrary to what we expected, we saw a dip in the latency around the same time for a device in Dunning. This is very likely a result of load-balancing applied by the ISP [1, 3] to address traffic demands during high network congestion. Devices `Logan Square 1`, `Logan Square 2`, `South Shore 1`, `South Shore 2` and `Lake View` were routed via a suboptimal traffic aggregation point, while `Dunning` was routed via a more optimal path. The second spike was not observed in *any* of the devices subscribed to the same ISP in Lake View, South Shore or Dunning. It was observed in 6 out of 10 devices in Logan Square, reaffirming that it was a result of congestion at a link close to or within Logan Square.

Result 2: Devices located very close to each other may not always exhibit similar last-mile latency spikes. In another scenario, we observe a co-occurrence of latency spikes for device `South Shore 2` for Atlanta and the ISP's last-mile (the spikes around 28 May 2022 00:00 am in Figure 3a). This suggests that there are instances where the ISP's last-mile access network may also be congested. Since the last-mile RTT includes the RTT for the home router, it is important to find whether these co-occurring spikes are the result of a congested router or they appear at the last-mile link. As a next step, we therefore look at the RTT for the same duration for a different device, `South Shore 3`, that was located only 53 meters from `South Shore 2` and shared the same last-mile IP address as `South Shore 2`. Figure 3b shows the slice of the last-mile RTT for `South Shore 3` for the same time duration as Figure 3a. We observe negligible similarities between the two devices in terms of co-occurring spikes. This implies that the spike for `South Shore 2` likely was a result of congestion at the home router.

Implications for sampling design. Our demonstrations suggest that latency bottlenecks may arise at different parts of the network. The reasons behind these bottlenecks may also vary with their location along the end-to-end network path. For example, the spikes that were observed for `South Shore 2` in Figure 3a may have appeared as a result of multiple users within a home consuming content from a streaming service at the same time. As a consequence, the home router started delaying the transfer of any new packets across the network, leading to a higher latency than normal. For Figures 1a and 2, higher latencies may have been a result of congestion at city or regional level traffic aggregation routers.

It is of course extremely challenging to anticipate the network-level faults that can lead to congestion. Yet, identifying devices and durations exhibiting elevated levels of congestion in real-time may have important implications towards constructing an effective data sample. For example, if common last-mile latency spikes are identified for a group of devices with a shared last-mile at the same time, sampling the results from multiple diagnostic tests for only one of these devices may suffice. This is because in such a scenario, the performance bottleneck is most likely the last-mile aggregation point, and sampling additional tests from multiple vantage points will provide redundant information. Extending this approach for all aggregation points until the ISP's edge would thus help identify hierarchies of variation of Internet performance and lead to datasets with reduced sampling bias.

5 Looking Ahead

Existing Internet performance datasets exhibit both spatial and temporal bias, yet the nature of that bias, how it may affect more general conclusions, and—most importantly—how to construct a generalizable sample across a geography, remain poorly understood. This paper does not yet present a sampling strategy or broader solution to this problem but instead draws attention to the problem by observing that even for a relatively stable and predictable Internet performance metric, latency, variability does not often correlate across nearby network vantage points.

Understanding Internet performance across space and time will ultimately require more coverage within a smaller geography, and a better understanding of the underlying infrastructure and how it varies across a geography. Given better information about infrastructure locations from ISPs, even within smaller geographies, we could use our approach to identify local factors which may (or may not) be contributing to performance anomalies that occur within neighborhoods. Future analysis could also isolate the performance of the last-mile link, which may be important to ultimately the extent to which the home network, access ISP, interconnect, or other portion of the infrastructure may ultimately be responsible for a performance anomaly. Finally our existing sample, which was not driven specifically by the need to understand the effects of infrastructure on performance variability, may ultimately need to be even more focused and local to help us better understand the relationship between geography and performance anomalies. Specifically, our deployments have been targeted towards revealing performance inequities across geographies [7]. While this data helped us compare performance across neighborhoods, they did not allow us to identify the variation of Internet performance along the smallest units of geography and time (e.g., multiple houses within the same city block).

Acknowledgments. This research was funded by NSF awards CNS-2223610, CNS-2213821, CNS-2319603, and CNS-2224687.

References

[1] James E Burns, Teunis J Ott, Anthony E Krzesinski, and Karen E Müller. "Path selection and bandwidth allocation in MPLS networks." In: *Performance Evaluation* 52.2-3 (2003), pp. 133–152.

[2] Romain Fontugne, Anant Shah, and Kenjiro Cho. "Persistent last-mile congestion: not so uncommon." In: *Proceedings of the ACM internet measurement conference.* 2020, pp. 420–427.

[3] Srikanth Kandula, Dina Katabi, Shantanu Sinha, and Arthur Berger. "Dynamic load balancing without packet reordering." In: *ACM SIGCOMM Computer Communication Review* 37.2 (2007), pp. 51–62.

[4] Kyle MacMillan, Tarun Mangla, James Saxon, Nicole P Marwell, and Nick Feamster. "A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7)." In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 7.1 (2023), pp. 1–26.

[5] *Measurement Lab.* URL: `https://www.measurementlab.net`.

[6] *Netrics.* URL: `https://github.com/internet-equity/nm-exp-active-netrics`.

[7] Ranya Sharma, Tarun Mangla, James Saxon, Marc Richardson, Nick Feamster, and Nicole P Marwell. "Benchmarks or Equity? A New Approach to Measuring Internet Performance." In: *A New Approach to Measuring Internet Performance (August 3, 2022)* (2022).

[8] Andrew S Tanenbaum. *Computer networks.* Pearson Education India, 2003.