# Reassessing the Constancy of End-to-End Internet Latency

Lily Davisson*, Joakim Jakovleski*, Nhiem Ngo*, Chau Pham*, and Joel Sommers
*Department of Computer Science, Colgate University, Hamilton NY USA*

{ldavisson,jjakovleski,nngo,cpham,jsommers}@colgate.edu

***Abstract*—A paper by Zhang *et al.* in 2001, "On the Constancy of Internet Path Properties" [1] examined the *constancy* of end-to-end packet loss, latency, and throughput using a modest set of hosts deployed in the Internet. In the time since that work, the Internet has changed dramatically, including the flattening of the autonomous system hierarchy and increased deployment of IPv6, among other developments. In this paper, we investigate the constancy of end-to-end Internet latency, revisiting findings of the earlier study. We use latency measurements from RIPE Atlas, choosing a set of 124 anchors with broad geographic distribution and drawn from 112 distinct autonomous systems. The earlier work of Zhang *et al.* relies on *changepoint detection* methods to identify mathematically constant time periods. We reimplement the two methods described in that earlier work and use them on the RIPE Atlas latency measurements. We also use a recently-published method (HMM-HDP) that has direct support in a RIPE Atlas API.**

**Comparing the three changepoint detection methods, we find that the two methods used in the earlier work may miss many changepoints caused by common level-shift events. Overall, we find that the recently proposed HMM-HDP method performs substantially better. Moreover, we find that delay spikes—as defined by the earlier work—are an order of magnitude less prevalent than 20 years ago. We also find that maximum change-free regions (CFRs) along paths that we observe in today's Internet are substantially longer than what was observed in 2001, regardless of the changepoint detection method used. In particular, the $50^{th}$ percentile maximum CFR was on the order of 30 minutes in the earlier study, but our analysis reveals it to be on the order of *3 days or longer*. Moreover, we find that CFR durations appear to have steadily increased over the past 5 years.**

***Index Terms*—Network latency, network delay, change-point detection**

## I. INTRODUCTION

Latency is a critical factor in Internet application performance and user quality of experience. On the web, increases in latency are associated with user abandonment and lower ad conversions [2], [3], and a number of services exist to monitor and diagnose latency [4], [5]. More broadly, there has been a great deal of research over the past decades on measuring and predicting Internet latency, studying its effects, and investigating mitigations, *e.g.*, [1], [6]–[25].

Understanding the *stability* or *constancy* of latency over time is also of critical importance, since transport protocol performance typically depends on estimates of round-trip time (RTT) and throughput can suffer in the face of dramatic shifts or variation in RTT [26]–[29]. A study by Zhang *et al.* $\approx 20$ years ago established characteristics of latency in terms of its *mathematical constancy*, *operational constancy*, and *predictive constancy* [1] using measurements collected from Paxson's NPD hosts [30]. In the time since that study, the Internet has changed in dramatic ways with the rise of IXPs and the flattening of the autonomous system hierarchy [31]–[37], massive increases in the number of home users connected through high-speed broadband [38]–[41], continued growth in bandwidth both in the core and at the edge [42], [43], and continued evolution in application popularity. In 2000, the web was still in its infancy and was the most popular application; today, streaming video constitutes well over 50% of all Internet traffic [41], [44]–[46].

In this paper, we evaluate the *constancy* of Internet latency. We specifically focus on the notions of *mathematical*, *operational*, and *predictive* constancy as described in the earlier influential work of Zhang *et al.* [1][1]. There are many differences in the Internet between that time and now, and our methods and the data we use are also rather different. In particular, we use latency measurements from the RIPE Atlas project [47], using anchor mesh measurements from 124 anchors distributed across 88 countries, 6 continents, and 112 autonomous systems. Each of the anchors we use is dual-stack, allowing us to evaluate constancy in both the IPv4 and IPv6 Internet; [1] only considers IPv4 since IPv6 was only lightly deployed at the time. Moreover, the work of Zhang *et al.* relied on two changepoint detection methods for segmenting time into periods of mathematical constancy, $CP_{Bootstrap}$ and $CP_{RankOrder}$. We reimplement those methods, and compare their efficacy with a recently proposed Hidden Markov Model-based approach (HMM-HDP), which is directly supported through a RIPE Atlas API [48]. In our comparison of the three changepoint detection methods, we find that each of the two methods used in the earlier work miss a number of changepoints caused by level-shift events. We also find that, to a lesser extent, they identify spurious changepoints in the midst of stable round-trip time measurements. Overall, we find that the recently proposed HMM-HDP method performs

---

*The first four authors contributed equally to this work.

[1] Zhang *et al.* considered constancy of packet loss, latency, and throughput; we restrict our study to characteristics of latency.

978-3-903176-40-9 ©2021 IFIP

substantially better.

In our detailed evaluation of the "steadiness" or constancy of Internet latency, we find significant differences between what is observed in today's Internet compared with the earlier findings. Specifically, in our analysis of the raw latency measurements, we find that delay spikes are an order of magnitude less prevalent than 20 years ago. In particular, latency measurements that are 10x the median latency or higher occur roughly every 1 in 10,000 measurements in the RIPE Atlas data we use, but occurred roughly every 1 in 1,000 in 2001. We examine the results of changepoint analysis on the raw latency time series and find that maximum change-free regions (CFRs) are substantially longer than what was observed in 2001, regardless of the changepoint detection method used. For example, whereas the authors of the earlier study found that "delay appears well-described as steady on time scales of 10–30 minutes", we find that delay can be characterized as steady on time scales of 15 minutes to several *hours*. Further, whereas the $50^{th}$ percentile maximum CFR was about 30 minutes in 2001, we observe it to be on the order of 3 *days* today. We also find that CFR durations appear to have steadily increased over the past 5 years, and that delay constancy on intra-continental paths varies substantially, with European and North American paths being the most stable for both IPv4 and IPv6. Because of differences in the underlying data sources used in the previous study and ours, one must be cautious in drawing too fine a comparison between our results and past findings. Nonetheless, our results and comparisons with the Zhang *et al.* study strongly suggest that Internet latency characteristics have changed dramatically over the past 20 years.

## II. RELATED WORK

Measuring and understanding characteristics of latency on Internet paths has been an object of study since the beginnings of the Internet [6], [10], [49] and its predecessor the ARPAnet [50]. These studies (*e.g.*, [49] in particular) observed level-shifts and other non-steady behavior over the course of a day. The measurement studies by Paxson in the late 1990s established a baseline for much of what was known about end-to-end delay in the early years of the commercial Internet [8], [30]. Since that time many works have sought to characterize delays in the Internet, *e.g.*, [15], [51]–[54]. Today, several ongoing projects collect and publish delay measurements across the Internet [47], [55], [56].

Many works have investigated active (probe-based) methods for measuring delay to improve over the ubiquitous `ping` tool which uses ICMP echo request/reply. For example, Baccelli *et al.* examined methods for unbiased probing [57], [58], Gummadi *et al.* developed a method for measuring delay between arbitrary DNS servers [9], and Pelsser *et al.* investigated the effect of load balancing on standard ICMP echo request/reply (ping) delay measurements [17]. Besides measurement of delay, a number of works have investigated measurement of *jitter* or *delay variation*, which is important in streaming application protocols, *e.g.*, [13], [23], [59]. Yet other works have used delay measurements (or relative delays between two or more packets) to infer Internet link capacities [60], available bandwidth [61], and network congestion [62]–[64].

The first work to look explicitly at the *constancy* of Internet delay along end-to-end paths was that of Zhang *et al.* [1], [65]. This work, of course, is the main inspiration for our study. The work by Zhang *et al.* developed methods for identifying time points (*changepoints*), between which Internet path performance properties can be considered constant. Recently, Mouchet *et al.* developed a new technique for identifying changepoints in a time series of Internet delay measurements [48], complementing a larger body of work on changepoint detection in time series (*e.g.*, see [66], [67]). In our work, we use the three changepoint detection techniques from each of these works [1], [48]. Although other works have examined the question of how end-to-end latency has changed in the Internet over time (*e.g.*, [54]), we are not aware of studies more recent than [1] that have looked at the constancy of Internet latency.

## III. DATA AND METHODS

In this section we describe the data used in our study and the methods used for analysis.

### A. Data

We obtained the latency measurements for this study from RIPE Atlas [47]. In particular, we use built-in anchor mesh ping measurements, which are performed periodically (every 4 minutes) among the full mesh of Atlas *anchors* [68]. Anchors are well-provisioned hosts located at academic institutions, IXPs, datacenters and the like. The primary reasons we used measurements from these hosts as opposed to the broader set of Atlas probes are that latency measurements have been observed to be more accurate [69] because of the more capable hardware on which anchors are deployed, and also that we desired a dual-stack set of hosts that has been observed to be stable and highly available within RIPE Atlas.

We selected 124 dual-stack anchors with broad geographic coverage—6 continents and 88 countries are represented[2]. Figure 1 shows a map of where anchors we used are located (using the fuzzed coordinates provided through RIPE Atlas APIs). The selected set of anchors also has good distribution across host networks: 112 distinct autonomous systems are represented. Anchors are well-connected to the Internet, and we recognize that—besides low coverage of the overall number of ASes and world geography—another limitation of our study is that latency characteristics observed through anchor mesh measurements may not be representative of what may be experienced by "ordinary" end hosts. We plan to expand our study in the future to consider additional endpoints to improve AS and geographic coverage. We note, however, that in the study by Zhang *et al.*, the authors used hosts within the NIMI platform [70], which were strongly biased toward

---

[2] The full list of anchors used in our study is available at https://www.cs.colgate.edu/~jsommers/data/probes_fqdns_tma21.csv.

deployments in the United States, and which were also well-connected (about half were deployed in academic institutions and the remainder at research organizations). Two data sets are described in [1] which were derived using 31 hosts or 49 hosts. Given the greater number of anchors we use, the number of distinct paths considered in our study is also much larger.

As with the earlier study, latency measurements in RIPE Atlas are round-trip time measurements, not one-way delays. Beyond that similarity, there are many differences between what was done earlier and how RIPE Atlas latency measurements are collected. In Zhang *et al.*, Poisson-modulated probing was used, with rates of 10 Hz or 20 Hz depending on which data set, and measurement between a pair of hosts was done continuously for one hour at a time. One data set used by those authors was collected in Winter 1999–2000 and the other collected about a year later. On RIPE Atlas, each latency measurement consists of three ICMP pings, with 4 minutes (240 seconds) separating each measurement. The data on which we focus for much of our study was collected in the two week period of January 1–January 15, 2020. We also consider data collected in January 1–15 from 2016–2021 for longitudinal analysis. Note that there are 15,120 individual RTT measurements between a given anchor pair (three probes, every four minutes, over two weeks) and that with 124 anchors, there are 7,626 distinct pairs[3]. We chose a two-week period to roughly match the quantity of measurements collected from [1]. We used both raw data available from RIPE Atlas as well as the results from running HMM-HDP analysis [48] on the same time period for each anchor pair considered in our study. In our results, we highlight differences in findings between the earlier study and ours, but because of the differences in context and data used in [1] and our work any direct comparisons are done with caution.

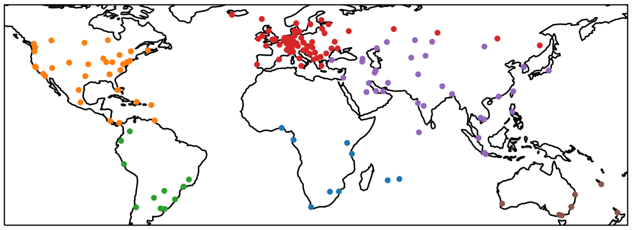

Fig. 1. Map of anchors in RIPE Atlas used in our study.

### *B. Methods*

Identifying changes between regions of *steady* or *constant* behavior was of critical importance in the methods employed by Zhang *et al.* [1]. In our work, we use three changepoint detection methods to analyze latency measurements for each distinct anchor pair. We re-implemented $CP_{Bootstrap}$ and $CP_{RankOrder}$, as described and used in [1]. We also used the HMM-HDP method described in Mouchet *et al.* [48] as available through a RIPE Atlas API. Figure 2 shows example timeseries plots with each of these three analyses for a path between an anchor in Barcelona, Spain and another in Frankfurt, Germany. One week of the trace is shown (7 Jan.–14 Jan. 2020). Vertical dashed lines in each time series indicate where a given changepoint detection algorithm has determined that a change has occurred. We observe for $CP_{Bootstrap}$ (top) and $CP_{RankOrder}$ (middle) that while some detected changepoints align with level-shifts in latency, others appear to be quite spurious. We expect level-shifts (change in the mean RTT) to represent one type of changepoint that any useful method should detect; another type of changepoint may be a change in RTT variance. We see that some clear level shifts are simply missed by these methods. We observe that the HMM-HDP method (bottom plot) identifies each level shift, while also identifying some (but many fewer) short-lived elevated latencies as changepoints.

We manually examined approximately 1,000 such plots, comparing the performance of the three changepoint detection methods. Identification of ground truth is challenging since there is no perfect automated way to detect level-shifts or similar changepoints. As Mouchet *et al.* [48] point out, it is quite natural for a human to be able to identify "obvious" changepoints, and in our qualitative manual comparison of these methods we relied on visual detection of level-shifts as our notion of ground truth. While this process certainly can admit human errors, it was overwhelmingly clear in our detailed manual study that the $CP_{Bootstrap}$ and $CP_{RankOrder}$ identified many more erroneous changepoints than did the HMM-HDP method. We also evaluated other off-the-shelf changepoint detection algorithms. Although we do not show results from those analyses, we found that performance varied; some were *worse* than the two methods used in Zhang *et al.*, *i.e.*, some "obvious" changepoints were missed, others spurious, and some were somewhat better. It is also important to note that Zhang *et al.* explicitly point out biases in the changepoint detection methods employed in their work. Based on our analyses, we generally focus on results obtained using the HMM-HDP method in the rest of the paper.

## IV. RESULTS

In this section, we present the results of our study. Where possible, we discuss our results in light of those found by Zhang *et al.* [1].

### *A. Delay Spikes*

We first investigate the presence of *delay spikes*, as was also done in prior work. As with Zhang *et al.* we computed the median RTT for each trace (each distinct anchor pair), then normalized each RTT measurement by the median. We show in Figure 3 a complementary distribution plot (notice that the y axis is log-scale) of these normalized RTTs. As with [1], we show a vertical reference line for a ratio of 10:1 (*i.e.*, a given RTT is 10 times higher than the median). We also show a horizontal reference line at a probability of $10^{-4}$.

---

[3] We used only one set of measurement data between a given anchor pair since the measurements are round-trip and our analysis yields the same results for each direction.

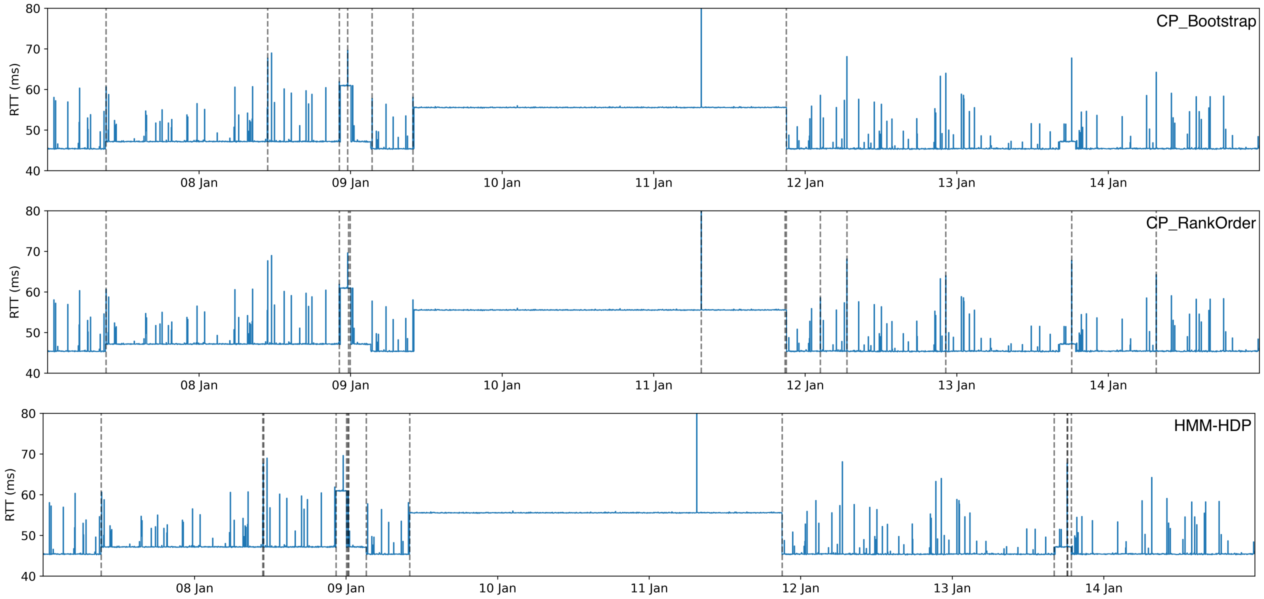

Fig. 2. Timeseries plots of latency from 7 Jan.–14 Jan. 2020 between Barcelona, Spain and Frankfurt, Germany (IPv4). Vertical dashed lines on each plot indicate the location at which a given changepoint detection algorithm determines that a change has occurred. The top plot: $CP_{Bootstrap}$; middle plot: $CP_{RankOrder}$; bottom plot: HMM-HDP [48].

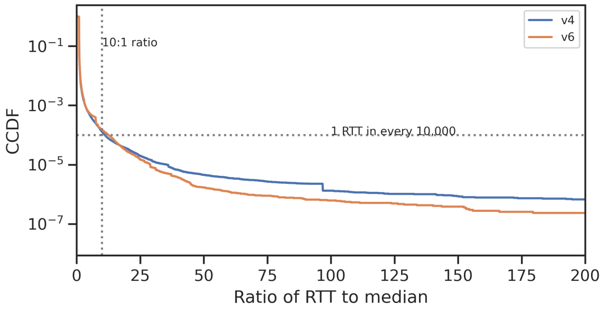

Fig. 3. Complementary distribution of the ratio of RTT samples to the median of their traces for IPv4 and IPv6; *cf.* Figure 8 in [1].

Interestingly, our horizontal reference is an *order of magnitude lower* than that shown in [1], indicating that such delay spikes are much less prevalent in the RIPE Atlas traces we use than what was observed in the earlier study. We also observe that there are somewhat fewer high RTT outliers in the IPv6 measurements than IPv4. Although our results show fewer RTT spikes than in [65], they are nonetheless consistent with observations by Padmanabhan *et al.* [25] and their observations of rare, extremely elevated, delay measurements.

### B. Constancy of RTT Distribution

In this section we analyze the results of applying changepoint detection algorithms to RIPE Atlas latency measurements between each distinct pair of anchors. The authors in the earlier study applied their changepoint detection methods to the *body* of the RTT distribution for each trace they used, examining the constancy of the RTT median as well as the constancy of the inter-quartile range (IQR). In our analysis, we looked at applying two of the changepoint detection methods we consider ($CP_{Bootstrap}$ and $CP_{RankOrder}$) in a similar manner, as well as using the raw trace data. Due to the low prevalence of spikes, we did not observe any significant difference in results whether spikes were filtered out or not. The authors of [1] note that the same was true in their analysis of operational constancy. Also, as noted above, since the HMM-HDP method [48] we use is available through a RIPE Atlas API we cannot force it to operate on, say, the IQR of a trace. As a consequence, the results we present in this section are derived from the applying the three changepoint methods on the raw/unfiltered trace data.

*1) Distribution of Change-free Region Durations:* In this section we examine the distribution of durations of *change-free regions* (CFRs), or time segment durations *between* detected changepoints, as a way to assess constancy of the distribution of RTTs along a path. We also investigate the distribution of the *maximum CFR* for the paths considered (in our case, for each distinct anchor pair), which was a focal point of the analysis done by Zhang *et al.*

In Figure 4, we show the distribution of CFR durations for each of the three changepoint methods, and for both IPv4 and IPv6; note that a comparable plot with the full distribution of CFR durations is not shown in [1] since they focused on the maximum CFR distribution. Notice that the x-axis is log-scale and that the units are seconds. There are several vertical dashed lines with different time durations shown for reference. We immediately see that the CFR durations computed using the HMM-HDP method are significantly shorter than those computed with $CP_{Bootstrap}$ and $CP_{RankOrder}$: the $50^{th}$ percentile CFR duration is *about 5.5 hours longer* using the methods from [1]. We also observe in Figure 4 that CFRs are slightly

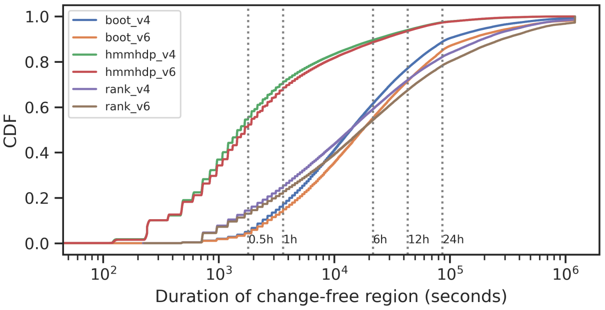

Fig. 4. CDFs of change-free durations for the three change-point detection methods we consider, and for IPv4 and IPv6 (first two weeks of January 2020). Note that the x-axis is log-scale number of seconds.

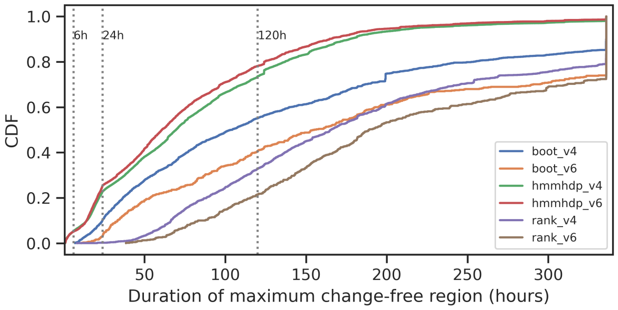

Fig. 5. CDF of the maximum duration of change-free region across anchor pairs considered (first two weeks of January 2020). Notice that the x-axis is in hours. *cf.* Fig. 9 in [1].

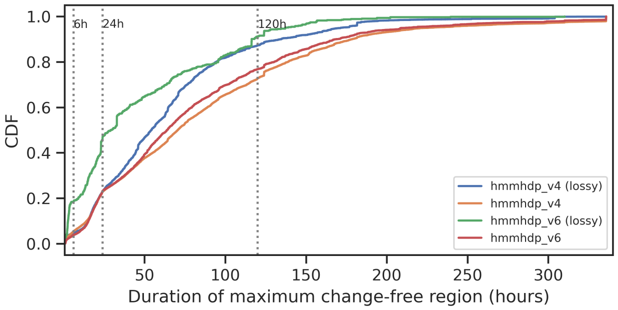

Fig. 6. CDFs of maximum duration of change-free region across anchor pairs considered, for the HMM-HDP method (first two weeks of January 2020). We show CDFs for *lossy* traces ($\geq$ 1% loss) and all traces. *cf.* Fig. 9 in [1].

lower (*e.g.*, there is a lower degree of constancy) in IPv4 vs. IPv6, but that there is a somewhat greater difference with the other two changepoint detection methods. Considering Figure 2 and our investigation of these changepoint detection methods, we expect that the true duration of constant regions between level shifts in RTT is close to the curves shown for the HMM-HDP method in Figure 4. In particular, in our detailed manual inspection of results from the two methods used in [1], we observed many instances of missed level shifts which, ideally, should be identified as changepoints, and which likely cause a skew in the distributions to the right. Overall, from Figure 4 we infer that delay-sensitive flows that persist longer than 15 minutes ($\approx 25^{th}$ percentile CFR duration for the HMM-HDP method) such as streaming video are likely to experience non-constant behavior such as level-shifts. Shorter-lived flows, however, are highly unlikely to experience such behavior. While the distribution of CFR durations does not say anything about the *magnitude* of the RTT change between one change-free region and another (the notion of *operational constancy*, discussed below, addresses this issue), it does indicate something important about the constancy of RTTs along a path.

In Figure 5 we focus on the distribution of the *maximum* CFR durations across the anchor pairs considered in our study, similar to [1]. Note that the x-axis unit is *hours* and that the maximum x-axis value represents the full two-week time period we consider. A much different picture emerges from this plot, specifically that Internet paths today exhibit a great deal of mathematical constancy compared with what was observed 20 years ago. In Figure 9 of [1], the authors show that the $50^{th}$ percentile maximum CFR duration is about 30 minutes or less, but our analysis indicates that the $50^{th}$ percentile maximum CFR duration is on the order of three or more *days*, depending on the changepoint detection method used. Recall that [1] did their changepoint analysis on the median and IQR of the latency distribution, which would likely result in a more stable view of latency even with the higher probe rate used in that work. We also observe an inflection point at 24h, particularly for changepoints computed using the HMM-HDP method, which may be due to daily traffic engineering adjustments, or possibly diurnal load fluctuations and congestion. Interestingly, we see that maximum CFR durations are slightly smaller with IPv6 than with IPv4 when using the HMM-HDP changepoint detection method.

In Figure 6, we show the distribution of maximum CFR durations for the HMM-HDP method and separately show curves for *lossy* anchor pairs for IPv4 and IPv6. We consider an anchor pair to be lossy if the raw trace shows there to be 1% packet loss or greater. The plot clearly shows that the lossy anchor pairs exhibit shorter maximum CFR durations. For IPv6, the difference between the lossy traces and the aggregate is quite stark, strongly suggesting that these anchor pairs are likely what causes the $50^{th}$ percentile maximum CFR duration to be shorter with IPv6 than IPv4 for the HMM-HDP method. It is also clear from this plot that, as with the earlier study, paths with moderate packet loss exhibit a much lower degree of steady behavior.

*2) Intra- and Inter-continental Characteristics:* In Figure 7 we show distributions of maximum CFR durations for the HMM-HDP method, separating anchor pairs residing within the same continent (intra-continental pairs) from those that cross a continental boundary (inter-continental pairs). The figure shows separate plots for IPv4 (top) and IPv6 (bottom), and individual curves are labeled with continent codes for the intra-continental pairs. Note the log scale on the x-axis. We see in the plot a fairly wide range of maximum CFRs for different continents. For IPv4 we see that Africa, Europe, and North America have the longest $50^{th}$ percentile maximum CFRs,

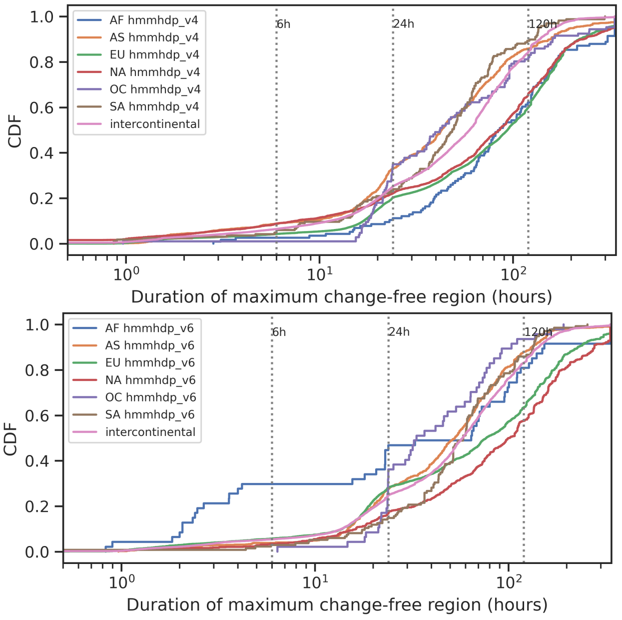

Fig. 7. CDFs of maximum duration of change-free region, separating anchor pairs by intra-continental (both anchors on the same continent) and inter-continental (anchors reside on two different continents). Top plot shows results for IPv4 and bottom plot shows results for IPv6. Note that the x-axis is on log scale.

while Asia, Oceania, and South America have the shortest (note that inter-continental maximum CFRs are in the middle of those groupings). For IPv6, however, no "natural groupings" emerge. Anchor pairs in Europe and North America still have the longest maximum CFRs in IPv6, and we also see that those maximum CFR durations are longer relative to IPv4. Note, however, that RTTs within some continents (*e.g.*, Europe) are on average naturally shorter than others (*e.g.*, Africa), so caution should be used in comparing intra-continental maximum CFRs. Interestingly, we observe that a number of anchor pairs in Oceania have a maximum CFR duration of 24h (notice the vertical segment aligned with 24h, which is quite significant in the IPv6 data). We plan to investigate these observed 24h behaviors in future work.

*3) Longitudinal characteristics:* Lastly, in Figure 8 we show distributions of maximum CFR durations for the HMM-HDP method for the first two weeks of January in the years 2016–2021. We note that we use a consistent set of anchors to construct each of these curves and that there were fewer total anchors in 2016. We performed our anchor selection such that we used as many long-lived anchors as possible, thus even in our 2016 analysis we use 58 anchors (already a larger number of hosts than used in [1]). We observe in the plots a general increase in maximum CFR duration over these years for both IPv4 and IPv6. For IPv4, we observe a reduction in maximum CFR between 2020 and 2021 which may be due to increased traffic volumes and delay during the COVID-19 pandemic [71], [72], but otherwise a general increase in maximum CFR. In future work, we plan to periodically update this plot in order to gauge the mathematical constancy of latency over time.

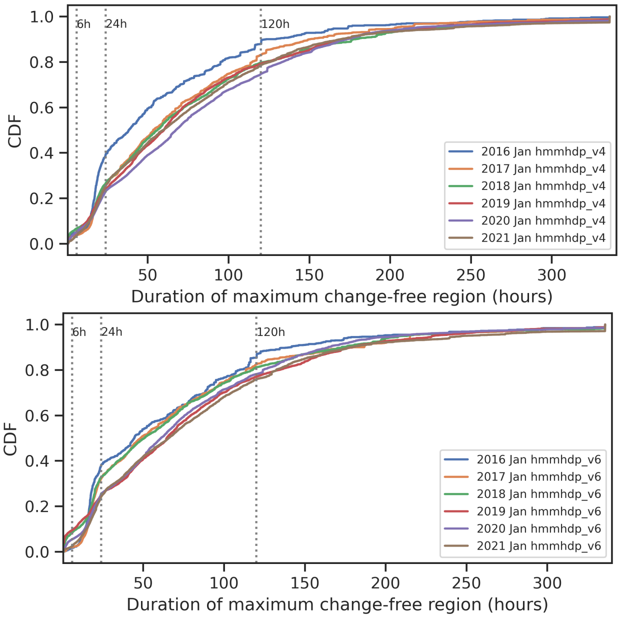

Fig. 8. CDFs of maximum duration of change-free region for years 2016–2021 (first two weeks of January). Top plot shows results for IPv4 and bottom plot shows results for IPv6.

### C. Operational Constancy

Latencies can be considered *operationally constant* if they remain within bounds that could be considered operationally equivalent. In the Zhang *et al.* study, the authors evaluated whether latencies remained within RTT categories of 0–100ms, 100–200ms, etc. They found that maximum CFRs for half the traces were under 10 minutes, and that 80% of traces had maximum CFRs under 20 minutes. In our analysis, we found that *more than 95%* of traces had maximum CFRs of 24 hours or longer. Even if we used a finer division of RTT ranges (e.g., 50 millisecond divisions) our results do not significantly change. Thus, we conclude that in today's Internet, although there can be considerable delay variation over short time scales as has been studied in prior work (*e.g.*, see [25], [38]), packet delay can be considered operationally steady on timescales of more than a day.

### D. Predictive Constancy

The third aspect of constancy we examine is predictive constancy. We used two families of predictors: a simple moving average with window sizes 2, 4, 8, 16, and 32, and an exponentially weighted moving average with $\alpha$= 0.5, 0.25, 0.125, and 0.01 (such that the moving average is $y_t = (1-\alpha)y_{t-1} + \alpha x_t$, where $x_t$ is the most recent sample and $y_0 = x_0$). We computed prediction error, as in [1], as:

$$\text{PredictionError} = E\left[\left|log\left(\frac{predicted}{actual}\right)\right|\right]$$

and generated CDFs to show the range of how well a given estimator performs.

Although we do not show detailed results due to space constraints, we found, as did Zhang *et al.*, that delay is *highly predictable*. Specifically, we found that the 95th percentile prediction error for all predictors used is less than 0.1, indicating high correspondence between the predicted and actual delay. Comparing with the earlier study, we observe that delay in the Internet today is *at least* as predictable as it was 20 years ago.

## V. SUMMARY AND CONCLUSIONS

In this paper we re-examine the constancy of end-to-end latency on Internet paths, which was first investigated by Zhang *et al.* [1] in 2001. We use anchor mesh latency measurements collected by the RIPE Atlas project and reimplement changepoint algorithms used the earlier study; we also use the recently proposed HMM-HDP time segmentation method for changepoint analysis [48]. We analyze latency spikes as well as distributional characteristics of time change-free regions. We perform specific analyses similar to Zhang *et al.* and compare findings of that work with what we observe in 2020, as well as longitudinally from 2016–2021. We find that delay spikes are an order of magnitude less prevalent than in the earlier study, and that Internet delays can be characterized as constant on timescales of 15 minutes to several hours, which is much longer than observed 20 years ago. We also find that maximum CFR durations are significantly longer than what was observed 20 years ago; the authors of [1] observed a $50^{th}$ percentile maximum CFR of about 30 minutes, but we observe it to be 3 days or longer. Moreover, we find that maximum CFR durations have steadily increased since 2016.

At the time of the study by Zhang *et al.*, the initial hype of the World-Wide Web was nearly passed. Although the Internet had existed for quite some time, only about 50% of adults in the United States were using the Internet in 2000, as compared with 90% or more today [40]. Moreover, very few people had home broadband Internet in 2000 in the United States, whereas roughly 75% of homes in the U.S. have broadband Internet today [40]. These specific observations change somewhat for different parts of the world, but it is also clear that over the past 20 years the Internet has become normalized as *critical infrastructure*. It should not, therefore, be surprising that the stability and constancy of Internet performance has improved over that time. Indeed, our results show that mathematical and operational constancy of Internet latency has improved a great deal in 20 years. In our future work, we plan to expand the set of RIPE Atlas anchors and probes used, and to consider additional public datasets in order to develop a more complete longitudinal picture of Internet latency. We hope that our work encourages other researchers to revisit seminal measurement studies as we collectively seek to understand Internet performance and its evolution over time.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their feedback and suggestions. This material is based upon work supported by the National Science Foundation under Grant No. CNS-1814537.

## REFERENCES

[1] Y. Zhang and N. Duffield, "On the constancy of Internet path properties," in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*. ACM, 2001, pp. 197–211.

[2] "Why performance matters," https://developers.google.com/web/fundamentals/performance/why-performance-matters/.

[3] H. Nam, K.-H. Kim, and H. Schulzrinne, "QoE matters more than QoS: Why people stop watching cat videos," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[4] "freshping: Reliable website uptime monitoring," freshworks.com/website-monitoring.

[5] "StatusCake: Website Uptime & Performance Monitoring," statuscake.com.

[6] J.-C. Bolot, "Characterizing end-to-end packet delay and loss in the Internet," *Journal of High Speed Networks*, vol. 2, no. 3, pp. 305–323, 1993.

[7] H. Balakrishnan, M. Stemm, S. Seshan, and R. H. Katz, "Analyzing stability in wide-area network performance," *ACM SIGMETRICS Performance Evaluation Review*, vol. 25, no. 1, pp. 2–12, 1997.

[8] V. Paxson, "End-to-end Internet Packet Dynamics," *IEEE/ACM transactions on networking*, vol. 7, no. 3, pp. 277–292, 1999.

[9] K. P. Gummadi, S. Saroiu, and S. D. Gribble, "King: Estimating latency between arbitrary Internet end hosts," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, 2002, pp. 5–18.

[10] D. Mills, "RFC0889: Internet delay experiments," 1983.

[11] H. V. Madhyastha, T. Anderson, A. Krishnamurthy, N. Spring, and A. Venkataramani, "A structural approach to latency prediction," in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, 2006, pp. 99–104.

[12] H. Pucha, Y. Zhang, Z. M. Mao, and Y. C. Hu, "Understanding network delay changes caused by routing events," *ACM SIGMETRICS performance evaluation review*, vol. 35, no. 1, pp. 73–84, 2007.

[13] T. Høiland-Jørgensen, B. Ahlgren, P. Hurtig, and A. Brunstrom, "Measuring latency variation in the Internet," in *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*, 2016, pp. 473–480.

[14] R. Durairajan, S. K. Mani, J. Sommers, and P. Barford, "Time's Forgotten: Using NTP to understand Internet latency," in *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, 2015, pp. 1–7.

[15] J. Chavula, N. Feamster, A. Bagula, and H. Suleman, "Quantifying the effects of circuitous routes on the latency of intra-Africa Internet traffic: a study of research and education networks," in *International Conference on e-Infrastructure and e-Services for Developing Countries*. Springer, 2014, pp. 64–73.

[16] R. Noordally, X. Nicolay, P. Anelli, R. Lorion, and P. U. Tournoux, "Analysis of Internet latency: the Reunion Island Case," in *Proceedings of the 12th Asian Internet Engineering Conference*, 2016, pp. 49–56.

[17] C. Pelsser, L. Cittadini, S. Vissicchio, and R. Bush, "From Paris to Tokyo: On the suitability of ping to measure latency," in *Proceedings of the 2013 conference on Internet measurement conference*, 2013, pp. 427–432.

[18] M. Claypool and D. Finkel, "The effects of latency on player performance in cloud-based games," in *2014 13th Annual Workshop on Network and Systems Support for Games*. IEEE, 2014, pp. 1–6.

[19] S. Sundaresan, W. De Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè, "Broadband Internet performance: a view from the gateway," *ACM SIGCOMM computer communication review*, vol. 41, no. 4, pp. 134–145, 2011.

[20] S. Roy and N. Feamster, "Characterizing correlated latency anomalies in broadband access networks," in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, 2013, pp. 525–526.

[21] B. Briscoe, A. Brunstrom, A. Petlund, D. Hayes, D. Ros, J. Tsang, S. Gjessing, G. Fairhurst, C. Griwodz, and M. Welzl, "Reducing Internet latency: A survey of techniques and their merits," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2149–2196, 2014.

[22] T. Flach, N. Dukkipati, A. Terzis, B. Raghavan, N. Cardwell, Y. Cheng, A. Jain, S. Hao, E. Katz-Bassett, and R. Govindan, "Reducing web latency: the virtue of gentle aggression," in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, 2013, pp. 159–170.

[23] J. Sommers, P. Barford, N. Duffield, and A. Ron, "Multiobjective monitoring for SLA compliance," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 652–665, 2009.

[24] A. Singla, B. Chandrasekaran, P. B. Godfrey, and B. Maggs, "The Internet at the speed of light," in *Proceedings of the 13th ACM Workshop on Hot Topics in Networks*, 2014, pp. 1–7.

[25] R. Padmanabhan, P. Owen, A. Schulman, and N. Spring, "Timeouts: Beware surprisingly high delay," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 303–316.

[26] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh, and V. Jacobson, "BBR: Congestion-based congestion control," *Queue*, vol. 14, no. 5, pp. 20–53, 2016.

[27] S. Shakkottai, N. Brownlee, A. Broido *et al.*, "The RTT distribution of TCP flows on the Internet and its impact on TCP based flow control," Cooperative Association for Internet Data Analysis (CAIDA), Tech. Rep., 2004.

[28] J. Martin, A. Nilsson, and I. Rhee, "Delay-based congestion avoidance for TCP," *IEEE/ACM Transactions on networking*, vol. 11, no. 3, pp. 356–369, 2003.

[29] M. Scharf, M. Necker, and B. Gloss, "The sensitivity of TCP to sudden delay variations in mobile networks," in *International Conference on Research in Networking*. Springer, 2004, pp. 76–87.

[30] V. Paxson, "Measurements and Analysis of End-to-end Internet Dynamics," Ph.D. dissertation, University of California at Berkeley, 1997.

[31] E. Gregori, A. Improta, L. Lenzini, and C. Orsini, "The impact of IXPs on the AS-level topology structure of the Internet," *Computer Communications*, vol. 34, no. 1, pp. 68–82, 2011.

[32] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and K. Claffy, "AS relationships, customer cones, and validation," in *Proceedings of the 2013 conference on Internet measurement conference*, 2013, pp. 243–256.

[33] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger, "There is more to IXPs than meets the eye," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 5, pp. 19–28, 2013.

[34] Y. Shavitt and U. Weinsberg, "Topological trends of Internet content providers," in *Proceedings of the Fourth Annual Workshop on Simplifying Complex Networks for Practitioners*, 2012, pp. 13–18.

[35] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "The flattening Internet topology: Natural evolution, unsightly barnacles or contrived collapse?" in *International Conference on Passive and Active Network Measurement*. Springer, 2008, pp. 1–10.

[36] Y.-C. Chiu, B. Schlinker, A. B. Radhakrishnan, E. Katz-Bassett, and R. Govindan, "Are we one hop away from a better Internet?" in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 523–529.

[37] A. Dhamdhere, C. Dovrolis *et al.*, "The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh," in *ACM SIGCOMM Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, no. 21, 2010.

[38] S. Sundaresan, W. De Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè, "Measuring home broadband performance," *Communications of the ACM*, vol. 55, no. 11, pp. 100–109, 2012.

[39] "National broadband plan," https://www.fcc.gov/general/national-broadband-plan.

[40] P. R. Center, "Internet/Broadband Fact Sheet," https://www.pewresearch.org/internet/fact-sheet/internet-broadband/, 2019.

[41] "Cisco Annual Internet Report (2018–2023) White Paper," https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[42] Y. Zhuang, J. Cappos, T. S. Rappaport, and R. McGeer, "Future internet bandwidth trends: An investigation on current and future disruptive technologies," *Secure Systems Lab, Dept. Comput. Sci. Eng., Polytech. Inst. New York Univ., New York, NY, USA, Tech. Rep. TR-CSE-2013-0411/01/2013*, 2013.

[43] "Exponential bandwidth growth and cost declines," https://www.networkworld.com/article/2187538/exponential-bandwidth-growth-and-cost-declines.html, April 2012.

[44] S. McCreary *et al.*, "Trends in wide area IP traffic patterns-A view from Ames Internet Exchange," in *ITC specialist seminar*, 2000.

[45] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, 2009, pp. 90–102.

[46] "2019 Internet Statistics, Trends & Data," https://dailywireless.org/internet/usage-statistics/.

[47] R. N. Staff, "Ripe Atlas: A Global Internet Measurement Network," *Internet Protocol Journal*, vol. 18, no. 3, 2015.

[48] M. Mouchet, S. Vaton, T. Chonavel, E. Aben, and J. Den Hertog, "Large-Scale Characterization and Segmentation of Internet Path Delays with Infinite HMMs," *IEEE Access*, vol. 8, pp. 16 771–16 784, 2020.

[49] K. C. Claffy, G. C. Polyzos, and H.-W. Braun, "Traffic characteristics of the T1 NSFNET backbone," in *IEEE INFOCOM'93 The Conference on Computer Communications, Proceedings*. IEEE, 1993, pp. 885–892.

[50] L. Kleinrock and W. E. Naylor, "On measured behavior of the ARPA network," in *Proceedings of the May 6-10, 1974, national computer conference and exposition*, 1974, pp. 767–780.

[51] H. Martin, A. McGregor, and J. Cleary, "Analysis of Internet delay times," in *Proceedings of Passive and Active Measurment Workshop (PAM)*, 2000.

[52] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and S. C. Diot, "Packet-level traffic measurements from the sprint ip backbone," *IEEE network*, vol. 17, no. 6, pp. 6–16, 2003.

[53] B.-Y. Choi, S. Moon, Z.-L. Zhang, K. Papagiannaki, and C. Diot, "Analysis of point-to-point packet delay in an operational network," *Computer networks*, vol. 51, no. 13, pp. 3812–3827, 2007.

[54] D. Lee, K. Cho, G. Iannaccone, and S. Moon, "Has Internet delay gotten better or worse?" in *Proceedings of the 5th International Conference on Future Internet Technologies*, 2010, pp. 51–54.

[55] "PingER: Ping End-to-End Reporting," https://www-iepm.slac.stanford.edu/pinger/.

[56] "Archipelago (Ark) Measurement Infrastructure," https://www.caida.org/projects/ark/.

[57] F. Baccelli, S. Machiraju, D. Veitch, and J. C. Bolot, "On optimal probing for delay and loss measurement," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, pp. 291–302.

[58] F. Baccelli, S. Machiraju, D. Veitch, and J. Bolot, "The role of PASTA in network measurement," *IEEE/ACM Transactions on Networking*, vol. 17, no. 4, pp. 1340–1353, 2009.

[59] C. Demichelis and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)," 2002.

[60] R. Kapoor, L.-J. Chen, L. Lao, M. Gerla, and M. Y. Sanadidi, "Capprobe: A simple and accurate capacity estimation technique," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 67–78, 2004.

[61] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, 2003, pp. 39–44.

[62] V. Bajpai, S. J. Eravuchira, and J. Schönwälder, "Dissecting last-mile latency characteristics," *ACM SIGCOMM Computer Communication Review*, vol. 47, no. 5, pp. 25–34, 2017.

[63] R. Fontugne, A. Shah, and K. Cho, "Persistent Last-mile Congestion: Not so Uncommon," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 420–427.

[64] M. Luckie, A. Dhamdhere, D. Clark, B. Huffaker, and K. Claffy, "Challenges in inferring internet interdomain congestion," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, 2014, pp. 15–22.

[65] Y. Zhang, V. Paxson, S. Shenker, and L. Breslau, "The stationarity of Internet path properties: Routing, loss, and throughput," ACIRI Technical report, Tech. Rep., 2000.

[66] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[67] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Processing*, vol. 167, p. 107299, 2020.

[68] "Ripe Atlas: Anchors," https://atlas.ripe.net/about/anchors.

[69] V. Bajpai, S. J. Eravuchira, and J. Schönwälder, "Lessons learned from using the RIPE Atlas platform for measurement research," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 3, pp. 35–42, 2015.

[70] V. Paxson, J. Mahdavi, A. Adams, and M. Mathis, "An architecture for large scale Internet measurement," *IEEE Communications Magazine*, vol. 36, no. 8, pp. 48–54, 1998.

[71] M. Candela, V. Luconi, and A. Vecchio, "Impact of the covid-19 pandemic on the internet latency: A large-scale study," *Computer Networks*, vol. 182, p. 107495, 2020.

[72] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez *et al.*, "The lockdown effect: Implications of the covid-19 pandemic on internet traffic," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 1–18.