An Investigation on Core Network Latency

Uwe Bauknecht
Universität Stuttgart, IKR
uwe.bauknecht@ikr.uni-stuttgart.de

Tobias Enderle
Universität Stuttgart, IKR
tobias.enderle@ikr.uni-stuttgart.de

***Abstract*—The growing adoption of 5G and cloud services places an increasing importance on the attainable point-to-point latency in Internet service provider networks. It directly impacts whether latency-critical services can be offered as well as the number and location of the corresponding data centers. In order to investigate the limits and variability of latency values in an actual core network, we collected round-trip time values published by a large North American Internet service provider spanning the duration of more than one year. We present a statistical analysis of this data set from which we infer a potential fiber topology. We use this topology to hypothesize on the efficacy of different means of reducing latency and determine their effect on the viability of low-latency services.**

***Index Terms*—Optical fiber networks, Quality of service, Routing, Telecommunication network topology, Wide area networks**

I. Introduction

A. Motivation

With the advent of access technologies like 5th Generation cellular communications (5G) or Fiber-to-the-x (FTTx) and an increasing availability and reliance on diverse cloud services in the business sector, high-quality network interconnections are an essential prerequisite to many future use cases from augmented reality applications to telemedicine and remote-operated vehicles. These use cases do not only require high data rates and availability of the underlying connection services. Low latency of the data transmission with little jitter is essential to their functionality, as well. Internet service providers (ISPs) are therefore increasingly interested in providing low-latency services at competitive prices. While purpose-built infrastructures with dedicated optical connections are only limited by the speed of light, they are also expensive to set up and maintain. We therefore investigate a non-purpose-built network by analyzing latency statistics of a regular IP network published by a large North American ISP, identifying common effects and determining their magnitude in order to estimate the viability of providing low-latency services.

Latency in core networks has been investigated from different angles. Some works focus on technological aspects of the optical domain [1], [2], while others concentrate on higher layers [3], [4]. Some have also included topological information in determining the attainable latency [2], [3], [5]. However, few works use actual data from commercial core networks. To the best of our knowledge, none employ data sets spanning more than a few months, while analyzing latency relative to the potential requirements of future services.

B. Sources of Latency in Networks

Latency in networks, i. e. the delay between sending data at a source and receiving it at the destination, is comprised of four major components.

Serialization delay is the time required to bring a data packet onto the transmission link. It depends on the link speed of the interface and the size of the data. For long-reach interfaces in core networks, where 100 Gbit/s is a common link speed, an Ethernet frame will incur about 100 ns of serialization delay.

Processing delay is incurred by packet processing in routers and switches. This includes the time to classify a received packet, making a forwarding decision and relaying it to the correct output port. This can take between hundreds of nanoseconds [6] and a few microseconds [7].

Queuing delay is the time a packet is stored in a buffer queue on a router. It results from contention when several packets are simultaneously destined for the same port. In the worst case, this constitutes an overload state when the rate of new arrivals exceeds the rate of departing packets. This delay is typically on the order of microseconds [7], but may reach hundreds of milliseconds during a substantial overload, due to the large buffers on backbone routers [8].

Propagation delay is the time a packet travels along a link. It depends on the transmission technology and its physical limits. For optical systems the fundamental limit is the speed of light in silica fiber. A length of 200 km incurs about 1 ms of delay, such that propagation delay dominates in large-scale core networks in the absence of overload events.

The length of fiber between two network nodes is typically longer than the distance "as the crow flies" by a factor of 1.2 to 1.5 [9]. This is mainly caused by the fact that fiber is deployed in conduits which are often dug adjacent to road, rail, power or gas lines. Furthermore, there is between 2 % to 10 % slack [10], [11] to account for thermal and tectonic effects as well as to provide spare fiber to aid repairs.

C. Organization of the Paper

In the following, we will first explain how the data set was collected. We will then present some basic properties of the collected delay values in Section III and infer a fiber topology for the network in Section IV. The topology is needed to argue about observed route changes and variable delay components in Section V. Furthermore, we present alternative routing approaches and their effects on the delay in Section VI before we conclude the paper.

II. Data Acquisition and Processing

A small number of large ISPs publish basic information about their networks' performance, including latency data between nodes in their backbone networks, which is typically given as a table of Round-Trip Times (RTTs) between the node pairs. We have surveyed this data of a large American ISP [12] for the duration of more than one year. According to the methodology information outlined in their supplementary resources [13], [14], the ISP continuously sends probes of UDP packet sequences to determine the RTT and aggregates the results, updating the published values every 15 minutes. We have sampled this data since 23 May 2019, until the ISP ceased publishing it in July of 2020.

The RTT table contains 25 nodes which yields 300 node pairs. However, for 41 of those pairs no RTT values have been published during our sampling period. All of these 41 node pairs either connect to the Madison or Indianapolis nodes. For the remaining 259 node pairs there exist gaps in which RTT values are missing. The majority of pairs (189) is missing less than 2 % of data, 58 are missing between 2 % and 7 %. For the remaining twelve node pairs the amount of missing data is between 16 % and 23 %. As above, many of them connect to Madison or Indianapolis, but also Chicago is involved. The final data set consists of 10,276,734 RTT values.

Since the published values appear to be rounded to full milliseconds, only propagation and queuing delays can be observable in the data set.

III. Round-Trip Time Characteristics

A. Long-Term Average of the Network-Wide RTT

To identify long-term trends we study the weekly average of the network-wide RTT over the whole sampling period. We consider the average for the whole data set including potential extreme values, as well as the average for a subset in which we ignore values exceeding the 99.99$^{\text{th}}$ percentile ($P_{99.99}$) for each node pair.

As can be seen in Fig. 1, the extreme values in the data set result in considerable peaks in the weekly network-wide average (points A, B and C). For points A and B those peaks can be traced back to an RTT peak which affects only a fraction of the node pairs and which is present for only one 15 minute interval. For point C the case is different. A detailed look at the data reveals that three node pairs suffer a sudden delay increase. For the following two days no RTT values have been provided for these pairs, hence it is unclear whether the delay increase was only temporary or of longer duration.

Until April 2020 a downwards trend is visible. This trend is interrupted in May 2020 and the RTT does not recover completely until the end of the sampling period.

B. Distribution and Variability of the RTT

In order to characterize the network delay in more detail, we consider the distribution of all measured RTT values. Fig. 2 shows the corresponding empirical complementary cumulative distribution function (CCDF) on a double logarithmic scale.

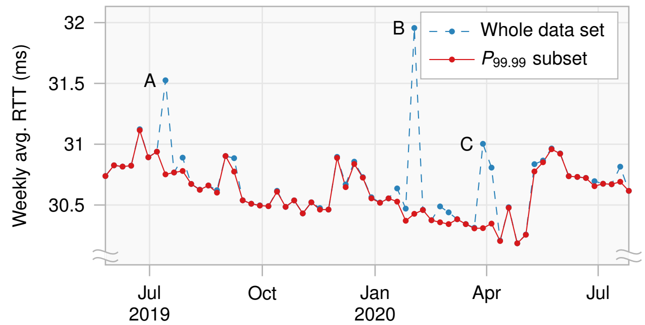


Fig. 1. Weekly averages of the network-wide RTT.

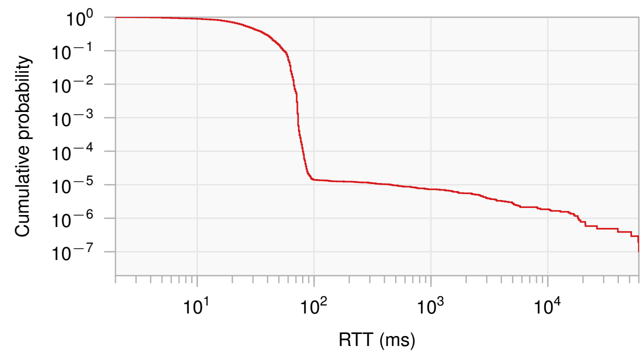


Fig. 2. Empirical CCDF of all measured RTT values.

90 % of the recorded RTT values are below 55 ms, which matches the geographic extent of the US (the approximate driving distance from Seattle to Orlando for example is 5000 km). Almost all measured values (99.9986 %) are in the region below 100 ms. However, starting from 100 ms, we can identify a large tail in the distribution with the highest recorded value being 62,071 ms. It is not entirely clear to us, whether such extremely high delays are solely caused by network effects or are measurement errors. Nevertheless, delays in the order of hundreds of milliseconds can certainly be attributed to congestion events in the network [14].

To get more insight, we study the variability of the delay per node pair. For each pair, we compute the difference of the 99.99$^{\text{th}}$ percentile and the minimum measured RTT and divide it by the minimum. By using the percentile instead of the maximum we reduce the influence of extremely high values which were possibly caused by measurement errors.

Fig. 3 shows box plots for this variability measure, grouped by the node pairs' minimum RTTs. The boxes show the first and third quartile and the median. The whiskers have a maximum length of 1.5 times the interquartile range but never exceed the data minimum or maximum, respectively, if they happen to be closer.

The diagram shows a trend towards less variability when the minimum RTT grows, i. e. when the path connecting the node pair gets longer. A possible explanation for this is that alternative paths for distant nodes might not differ much in length. On the other hand, for nearby nodes, which might be

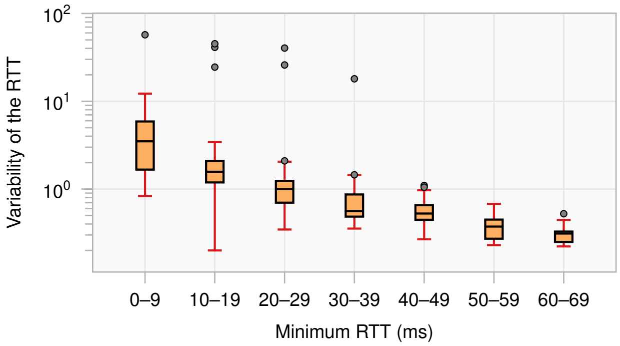


Fig. 3. Variability of the RTT.

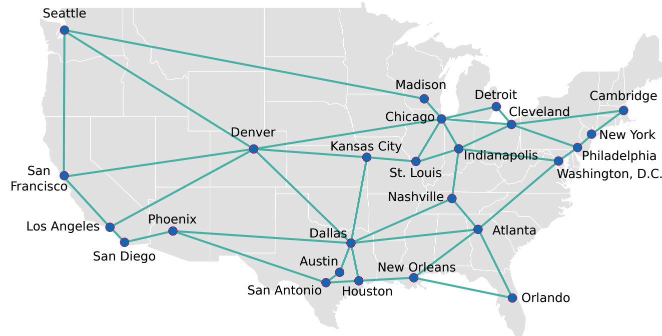


Fig. 4. Inferred fiber topology of the network.

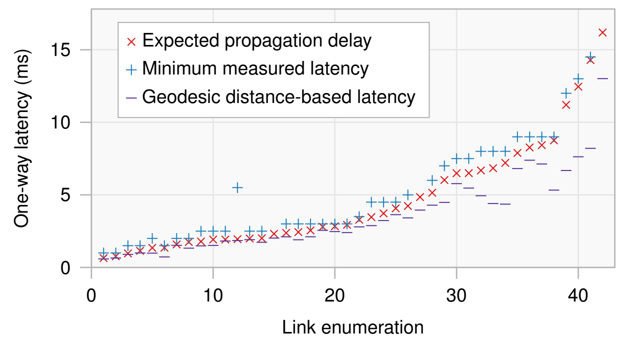


Fig. 5. Latency values for each link of the topology presented in Fig. 4.

connected using very few hops, already the second shortest path often results in a considerable detour. Except for some single node pairs, the variability is well below 10 and for higher minimum delays even below 1, which indicates that the RTT is quite stable.

IV. Fiber Network Topology

The knowledge of paths in the network helps to understand latency effects. While several researchers have already done excellent work on determining plausible topologies for this network, their data was either not available to us [5] for legal reasons or was focused on IP adjacencies [15], rather than fiber links. We have therefore inferred a plausible fiber topology based on our collected data, the publicly available sources used in [5], older published versions of the network [16], as well as road and train network maps.

We determined the lowest RTT observed for the cities in our data and divided this value by 2 in order to obtain an approximation to the one-way latency, assuming symmetric routing to be likely for the lowest observed values. We correlated these values with latency values obtained for road and train tracks, by multiplying their lengths with a value of 4.8985 µs/km and including 2 % of slack. We started from the lowest latencies, creating new links whenever the recorded value was too low, to allow a hop via one of the other cities. The resulting topology is shown in Fig. 4, while Fig. 5 illustrates the difference between the latency minima obtained from the data set and the expected propagation delay of the inferred links. We have also included latency values based on the geodesic distance between the cities, i. e. the great-circle distance on the earth sphere or “as the crow flies”, for comparison.

The nodes at Madison and Indianapolis seem to have a special role, since the published RTT information for these cities only covers few node pairs. As a consequence, links 15, 27 and 42 in Fig. 5, which connect to these nodes, do not show any measurement value. Regardless, the fiber conduits between Chicago and Seattle likely pass through Madison via Minneapolis, while the connection between St. Louis and Washington, D.C. is likely to traverse Indianapolis. The latter connection is especially interesting as its minimum latency implies the presence of fiber conduits between St. Louis and Indianapolis. However, this conduit is not used by the connection between these cities according to the measurements, which suggest that St. Louis – Indianapolis is routed via Chicago, explaining the outlier at index 12. It is also noteworthy that some longer fiber connections, especially at Denver and Dallas with indices above 30, seem to incur larger delays than could be anticipated based on road and geodesic distance, but since these are also among the longest links while traversing sparsely populated areas, the margin for error is certainly larger. Overall, the average difference between our estimate and the measurement without the outlier at index 12 is 590 µs with a maximum difference for the connection between Dallas and Nashville (index 32) at 1.32 ms. We use this inferred topology in our further analyses in order to separate effects on links from effects on multi-hop paths.

V. Aspects of Latency Components

A. Propagation Delay and Path Changes

As stated in Section II, only propagation and queuing delays will be discernible in the data. Following the formation of the topology, we further investigate the latency values regarding the likelihood of path changes which lead to large, but stable changes. Fig. 6 shows the minimum, average, maximum and the 99^{th} percentile of a subset of the recorded RTT data set, which includes only values corresponding to links of the inferred topology. For links 11, 27 and 42, no values are available, while for 11 other links, the maximum is not shown, since

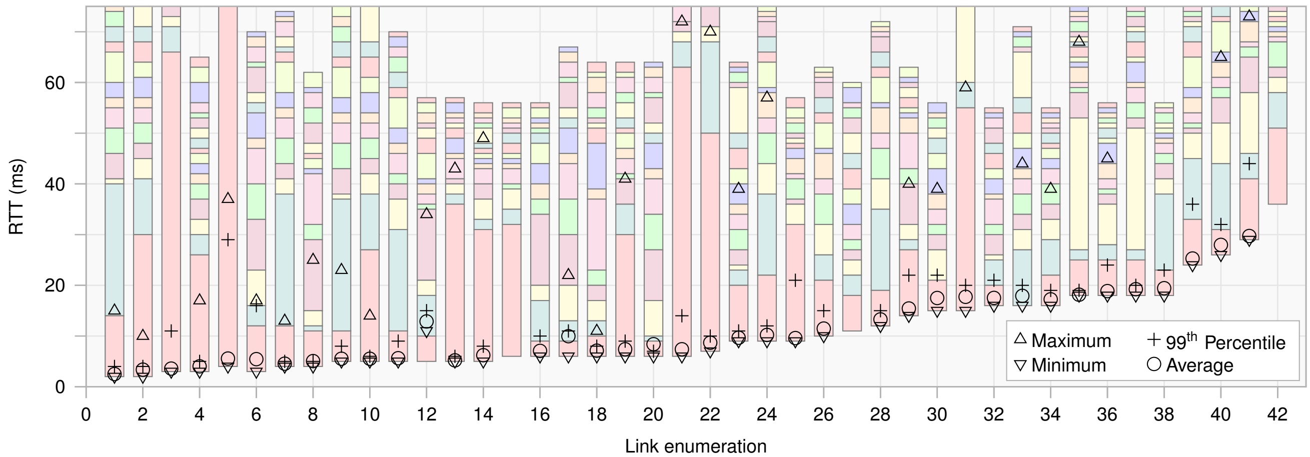


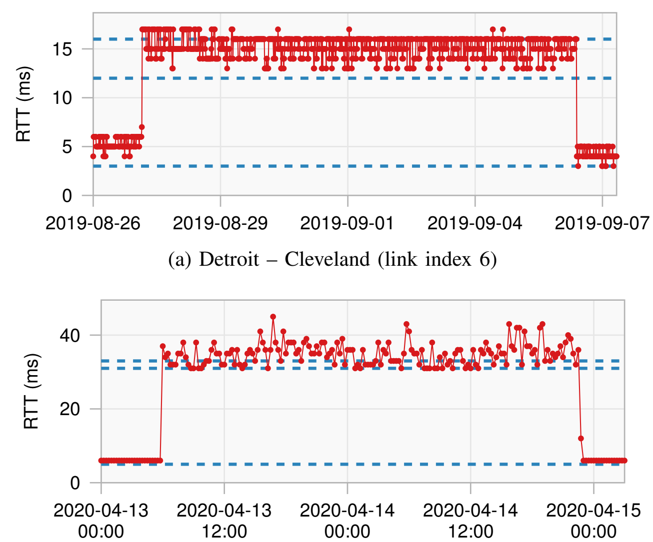
Fig. 6. RTT statistics for neighboring nodes.

these values are beyond 0.6 s, also including the outlier of more than one minute identified in Section III-B.

In order to provide an indication whether the variation of the observed values could be related to routing changes, leading to different propagation delays, we determined the minimum RTT incurred on each of the 20 shortest paths for each node pair, based on the assumption that each link incurs exactly the minimum RTT observed in the data set. We present these values as 20 stacked bar plots, where the red bars directly above the minima reflect the range in which a measured RTT value can only have occurred on the shortest path, possibly including queuing delays. The blue bars directly on top of these red bars represent the minimum RTT for the 2nd shortest path, such that a value in this range might be caused by either the propagation delay on that path or by a substantial amount of queuing delay on shorter paths. Values within upper bars therefore do not guarantee that a longer path has been used.

For all of these single-hop connections, the average RTT is fairly close to the minimum and remains within the lowest red bars, such that traffic is routed on the shortest path. There are only three connections for which the average lies outside the red bars. At index 12, which corresponds to the outlier in Fig. 5, the direct connection is never used, such that minimum, average and the 99th percentile are in the range for the second shortest path. Index 33 also shows the three marked values within the blue bar, but this is a special case, since the shortest and second shortest paths are equal in length due to the course rounding of the latency values in the data set, such that no red bar is shown.

At index 17, which is the connection between Detroit and Chicago, the difference between minimum and average RTT is the largest, while the 99th percentile and average values are very similar. In fact, starting from 24 September 2019, the RTT is strictly above the RTT of the second shortest path, which indicates a route change rather than fluctuations due to congestion. For 32 of the 42 connections, the 99th percentiles remain within the lower red bars. For the remaining 10 connections, this value is never larger than what is possible



(a) Detroit – Cleveland (link index 6)

(b) Nashville – Atlanta (link index 14)

Fig. 7. Sections of RTT values that suggest persistent route changes. The dashed lines indicate the three shortest paths based on the inferred topology.

on the second or third shortest paths, which might suggest that path changes are indeed a relatively rare occurrence. Nevertheless, Fig. 7 exemplarily presents two sections of RTT values for the links Detroit – Cleveland (index 6) and Nashville – Atlanta (index 14), for which persistent route changes lasting hours or even days are almost certain. The dashed lines indicate the three shortest paths based on the inferred topology.

When considering the maximum values in Fig. 6 however, a large fraction of connections exhibits temporary latency excursions, which are significant enough to indicate changes to drastically longer routes or large congestion events. All of these maxima above 75 ms, however, are only single values within the data. While any data point represents an interval of 15 minutes, such that a change to a longer route and back again can be accomplished, it seems more likely that such transient values are caused by congestion events or other malfunctions.

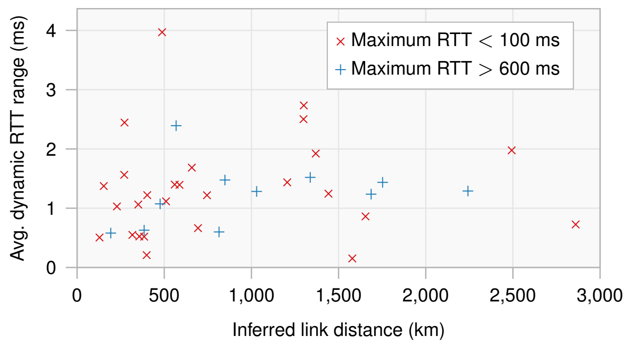


Fig. 8. Average dynamic RTT range for different connections using direct links. No maximum values were recorded between 100 and 600 ms.

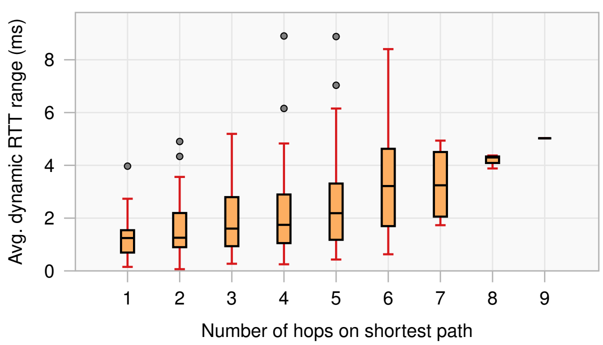


Fig. 9. Average dynamic RTT range for paths of different hop counts.

B. Queuing Delay and Variability

Further investigating the difference between the average and the minimum observed RTT, which we will refer to as the *average dynamic range* of the RTT, we plot its values over the length of the direct connections in Fig. 8. All values lie within a range of 0.15 ms to 3.97 ms. This is well in line with what can be expected as queuing delay, which is the only dynamic component of latency that does not depend on hardware parameters. Further supporting the assumption that the average dynamic range is the result of queuing events is the fact that it is largely independent of the link distance. In Fig. 8, we also highlight, which of these values include the impact of the previously mentioned spikes in the maximum RTT (max. RTT above 600 ms), indicating that these short-term latency excursions cause little skew to the average values.

Fig. 9 explores how the average dynamic range is affected by the number of hops on the shortest path. The values from Fig. 8, in which the shortest path requires only one link, are represented as the first box plot. Paths of multiple hops show significantly higher median values for the average dynamic range, as queuing delays from every hop can accumulate and increase the overall RTT. This effect also contributes to a growing variability for a larger number of hops, which is only diminished for paths of 7 hops and beyond, because the number of samples is decreasing rapidly due to the size of the topology. Regardless, the increasing median values and variability may become prohibitive to low-latency services.

VI. Means of Reducing Latency

The viability of offering low-latency services on point-to-point connections in the given network topology is subject to several factors. The geodesic distance between a pair of cities and the speed of light in fiber provide a firm lower bound to the lowest latency that can be achieved. As shown in Fig. 5, fiber is typically not deployed on the shortest route possible, but on the cheapest among the shortest routes, with trenches following preexisting structures such as roads or train tracks.

Fig. 10 shows an enumeration of all 259 node pairs, for which data could be recorded, ordered by the minimum RTT within the data set. The lowest graph (magenta) shows the RTT if all links of the topology had a length corresponding to the geodesic distance. This represents an unrealistic lower bound, which can be approached by very expensive, more direct fiber deployment. The next graph from the bottom (purple) gives a hypothetical RTT based only on propagation delay of the inferred link distances, which most likely underestimates the amount of fiber deployed by a small margin.

We also show the average (orange) and the 99.99th percentile (red). While the average remains remarkably close to the propagation delay, suggesting ranges in which low-latency services may indeed be viable, the percentile shows dramatic latency excursions, with outliers above 100 ms, which are caused by relatively few, short-term events. However, given that low-latency services require guarantees around similar percentiles, such deviations should be avoided.

This can be achieved by employing dedicated hardware, bypassing the IP router entirely, or by Quality of Service (QoS) prioritization. We emulate values for the bypass solution by taking the minimum RTT value recorded and subtracting a hypothetical processing delay of 20 µs for each intermediate hop on the shortest path, which yields the black graph. While this approach is less expensive than deploying new fiber, it also results in a smaller impact. QoS prioritization is a software feature commonly found on backbone routers, where packets of low-latency services are scheduled ahead of any packets of regular, best-effort traffic. Depending on the amount of QoS traffic, however, there may exist contention between different flows of already prioritized low-latency traffic. To determine an optimistic estimation for this option, we take the 5th percentile of the average dynamic range of the single-hop paths and add this value multiplied by the number of hops to the minimum value recorded and plot the results in blue. We also provide a pessimistic version by repeating the same process for the 99th percentile of the average dynamic range (green).

In Fig. 11 we hypothesize on the potential to offer different low-latency services based on the aforementioned approaches for RTT reduction. A service of 30 ms or less can only be offered for less than 20 % of node pairs based on the 99.99th percentile. Potentially more lucrative services of 5 ms or less are impossible to deliver. In contrast to this, the average shows that low-latency services are in fact realizable on more than half of the connections. However, it takes either QoS prioritization or router bypassing to achieve services of 2 ms, which are

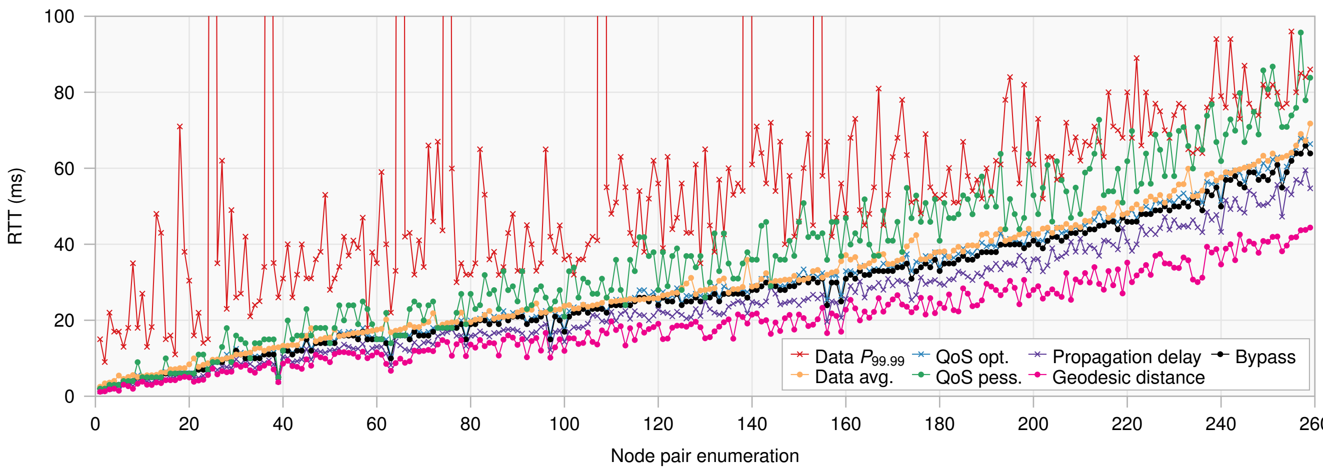


Fig. 10. RTT for various approaches.

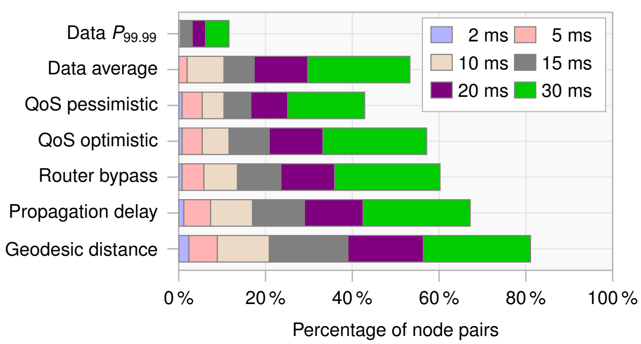


Fig. 11. Fraction of latency limits for node pairs.

potentially much more lucrative. We have also included the theoretical maximum of services by including the limits based on the geodesic distances and pure propagation delay.

VII. Conclusion

Based on the analysis of the acquired data we conclude that the average connection is not only relatively stable, but also that a trend to reduce the average latency is visible. Low average delays on many connections suggest that there is viability for offering low-latency services, but some latency excursions exist, such that they have to be bounded by technical means or explicitly considered in contractual agreements. Furthermore, our analysis based on the inferred topology has shown that QoS prioritization or router bypassing double the number of connections where an RTT of 5 ms can be achieved and even allow for 2 ms for some shorter links. While more general aspects of these findings most likely apply to other North American ISPs, a more detailed comparison of geographically diverse networks is considered for future work.

References

[1] V. Bobrovs, S. Spolitis, and G. Ivanovs, "Latency causes and reduction in optical metro networks," in *Photonics West - Optoelectronic Materials and Devices*, 2013.

[2] I. N. Bozkurt, W. Aqeel, D. Bhattacherjee, B. Chandrasekaran, P. B. Godfrey, G. Laughlin, B. M. Maggs, and A. Singla, "Dissecting latency in the Internet's fiber infrastructure," *arXiv:1811.10737 [cs.NI]*, 2018.

[3] 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.

[4] H. Pucha, Y. Zhang, Z. M. Mao, and Y. C. Hu, "Understanding network delay changes caused by routing events," *SIGMETRICS Perform. Eval. Rev.*, vol. 35, no. 1, p. 7384, Jun. 2007. [Online]. Available: https://doi.org/10.1145/1269899.1254891

[5] R. Durairajan, P. Barford, J. Sommers, and W. Willinger, "Intertubes: A study of the US long-haul fiber-optic infrastructure," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, p. 565578, Aug. 2015.

[6] "Low-latency data center and financial networking," Arista. [Online]. Available: https://www.arista.com/assets/data/pdf/JointPapers/Arista-Infinera-Low-Latency-Application-Brief.pdf

[7] V. Joseph and B. Chapman, "Chapter 6 - QoS service assurance," in *Deploying QoS for Cisco IP and Next Generation Networks*, V. Joseph and B. Chapman, Eds. Boston: Morgan Kaufmann, 2009.

[8] *Cisco Network Convergence System NCS 5500 Modular Platform Architecture*, Cisco, 2017. [Online]. Available: https://www.cisco.com/c/dam/en/us/products/collateral/routers/network-convergence-system-5500-series/ncs5500-modular-platform-architecture-white-paper.pdf

[9] *End-to-end error performance parameters and objectives for international, constant bit-rate digital paths and connections*, ITU Telecommunication Standardization Sector ITU-T Recommendation G.826, Dec. 2002.

[10] J. Botha, "OSP fiber optics civil works guide," The Fiber Optic Association, Inc. (FOA), 2015. [Online]. Available: https://www.thefoa.org/tech/ref/1pstandards/OSP%20Civil%20Works%20Guide-FOA.pdf

[11] *Criteria for optical fibre cable installation with minimal existing infrastructure*, ITU Telecommunication Standardization Sector ITU-T Recommendation L.163, Nov. 2018.

[12] AT&T. Network latency - global IP network - AT&T. [Online]. Available: http://ipnetwork.bgtmo.ip.att.net/pws/network_delay.html (Last accessed: 21 July 2020).

[13] ——. (2003) The quality of Internet service: AT&T's global IP network performance measurements. [Online]. Available: http://ipnetwork.bgtmo.ip.att.net/pws/paper.pdf (Last accessed: 21 July 2020).

[14] L. Ciavattone, A. Morton, and G. Ramachandran, "Standardized active measurements on a tier 1 IP backbone," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 90–97, 2003.

[15] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with rocketfuel," *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, pp. 2–16, 2004.

[16] AT&T. (2001) AT&T OC-48 fiber network. [Online]. Available: http://www.ipservices.att.com/backbone/bbonemap.cfm (Last accessed: 18 August 2002).