RFC 9318
IAB Workshop Report: Measuring Network Quality for End-Users

Abstract

The Measuring Network Quality for End-Users workshop was held virtually by the Internet Architecture Board (IAB) on September 14-16, 2021. This report summarizes the workshop, the topics discussed, and some preliminary conclusions drawn at the end of the workshop.

Note that this document is a report on the proceedings of the workshop. The views and positions documented in this report are those of the workshop participants and do not necessarily reflect IAB views and positions.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Architecture Board (IAB) and represents information that the IAB has deemed valuable to provide for permanent record. It represents the consensus of the Internet Architecture Board (IAB). Documents approved for publication by the IAB are not candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9318.

Copyright Notice

Table of Contents

1. Introduction

The Internet Architecture Board (IAB) holds occasional workshops designed to consider long-term issues and strategies for the Internet, and to suggest future directions for the Internet architecture. This long-term planning function of the IAB is complementary to the ongoing engineering efforts performed by working groups of the Internet Engineering Task Force (IETF).

The Measuring Network Quality for End-Users workshop [WORKSHOP] was held virtually by the Internet Architecture Board (IAB) on September 14-16, 2021. This report summarizes the workshop, the topics discussed, and some preliminary conclusions drawn at the end of the workshop.

1.1. Problem Space

The Internet in 2021 is quite different from what it was 10 years ago. Today, it is a crucial part of everyone's daily life. People use the Internet for their social life, for their daily jobs, for routine shopping, and for keeping up with major events. An increasing number of people can access a

gigabit connection, which would be hard to imagine a decade ago. Additionally, thanks to improvements in security, people trust the Internet for financial banking transactions, purchasing goods, and everyday bill payments.

At the same time, some aspects of the end-user experience have not improved as much. Many users have typical connection latencies that remain at decade-old levels. Despite significant reliability improvements in data center environments, end users also still often see interruptions in service. Despite algorithmic advances in the field of control theory, one still finds that the queuing delays in the last-mile equipment exceeds the accumulated transit delays. Transport improvements, such as QUIC, Multipath TCP, and TCP Fast Open, are still not fully supported in some networks. Likewise, various advances in the security and privacy of user data are not widely supported, such as encrypted DNS to the local resolver.

Some of the major factors behind this lack of progress is the popular perception that throughput is often the sole measure of the quality of Internet connectivity. With such a narrow focus, the Measuring Network Quality for End-Users workshop aimed to discuss various topics:

- What is user latency under typical working conditions?
- How reliable is connectivity across longer time periods?
- Do networks allow the use of a broad range of protocols?
- What services can be run by network clients?
- What kind of IPv4, NAT, or IPv6 connectivity is offered, and are there firewalls?
- What security mechanisms are available for local services, such as DNS?
- To what degree are the privacy, confidentiality, integrity, and authenticity of user communications guarded?
- Improving these aspects of network quality will likely depend on measuring and exposing metrics in a meaningful way to all involved parties, including to end users. Such measurement and exposure of the right metrics will allow service providers and network operators to concentrate focus on their users' experience and will simultaneously empower users to choose the Internet Service Providers (ISPs) that can deliver the best experience based on their needs.
- What are the fundamental properties of a network that contributes to a good user experience?
- What metrics quantify these properties, and how can we collect such metrics in a practical way?
- What are the best practices for interpreting those metrics and incorporating them in a decision-making process?
- What are the best ways to communicate these properties to service providers and network operators?
- How can these metrics be displayed to users in a meaningful way?

2. Workshop Agenda

The Measuring Network Quality for End-Users workshop was divided into the following main topic areas; see further discussion in Sections 4 and 5:

- Introduction overviews and a keynote by Vint Cerf
- Metrics considerations
- Cross-layer considerations
- Synthesis
- Group conclusions

3. Position Papers

The following position papers were received for consideration by the workshop attendees. The workshop's web page [WORKSHOP] contains archives of the papers, presentations, and recorded videos.

- Ahmed Aldabbagh. "Regulatory perspective on measuring network quality for end users" [Aldabbagh2021]
- Al Morton. "Dream-Pipe or Pipe-Dream: What Do Users Want (and how can we assure it)?" [Morton2021]
- Alexander Kozlov. "The 2021 National Internet Segment Reliability Research"
- Anna Brunstrom. "Measuring network quality - the MONROE experience"
- Bob Briscoe, Greg White, Vidhi Goel, and Koen De Schepper. "A Single Common Metric to Characterize Varying Packet Delay" [Briscoe2021]
- Brandon Schlinker. "Internet Performance from Facebook's Edge" [Schlinker2019]
- Christoph Paasch, Kristen McIntyre, Randall Meyer, Stuart Cheshire, and Omer Shapira. "An end-user approach to the Internet Score" [McIntyre2021]
- Christoph Paasch, Randall Meyer, Stuart Cheshire, and Omer Shapira. "Responsiveness under Working Conditions" [Paasch2021]
- Dave Reed and Levi Perigo. "Measuring ISP Performance in Broadband America: A Study of Latency Under Load" [Reed2021]
- Eve M. Schooler and Rick Taylor. "Non-traditional Network Metrics"
- Gino Dion. "Focusing on latency, not throughput, to provide better internet experience and network quality" [Dion2021]
- Gregory Mirsky, Xiao Min, Gyan Mishra, and Liuyan Han. "The error performance metric in a packet-switched network" [Mirsky2021]
- Jana Iyengar. "The Internet Exists In Its Use" [Iyengar2021]
- Jari Arkko and Mirja Kuehlewind. "Observability is needed to improve network quality" [Arkko2021]
- Joachim Fabini. "Network Quality from an End User Perspective" [Fabini2021]

- Jonathan Foulkes. "Metrics helpful in assessing Internet Quality" [Foulkes2021]

- Kalevi Kilkki and Benajamin Finley. "In Search of Lost QoS" [Kilkki2021]

- Karthik Sundaresan, Greg White, and Steve Glennon. "Latency Measurement: What is latency and how do we measure it?"

- Keith Winstein. "Five Observations on Measuring Network Quality for Users of Real-Time Media Applications"

- Ken Kerpez, Jinous Shafiei, John Cioffi, Pete Chow, and Djamel Bousaber. "Wi-Fi and Broadband Data" [Kerpez2021]

- Kenjiro Cho. "Access Network Quality as Fitness for Purpose"

- Koen De Schepper, Olivier Tilmans, and Gino Dion. "Challenges and opportunities of hardware support for Low Queuing Latency without Packet Loss" [DeSchepper2021]

- Kyle MacMillian and Nick Feamster. "Beyond Speed Test: Measuring Latency Under Load Across Different Speed Tiers" [MacMillian2021]

- Lucas Pardue and Sreeni Tellakula. "Lower-layer performance not indicative of upper-layer success" [Pardue2021]

- Matt Mathis. "Preliminary Longitudinal Study of Internet Responsiveness" [Mathis2021]

- Michael Welzl. "A Case for Long-Term Statistics" [Welzl2021]

- Mikhail Liubogoshchev. "Cross-layer Cooperation for Better Network Service" [Liubogoshchev2021]

- Mingrui Zhang, Vidhi Goel, and Lisong Xu. "User-Perceived Latency to Measure CCAs" [Zhang2021]

- Neil Davies and Peter Thompson. "Measuring Network Impact on Application Outcomes Using Quality Attenuation" [Davies2021]

- Olivier Bonaventure and Francois Michel. "Packet delivery time as a tie-breaker for assessing Wi-Fi access points" [Michel2021]

- Pedro Casas. "10 Years of Internet-QoE Measurements. Video, Cloud, Conferencing, Web and Apps. What do we Need from the Network Side?" [Casas2021]

- Praveen Balasubramanian. "Transport Layer Statistics for Network Quality" [Balasubramanian2021]

- Rajat Ghai. "Using TCP Connect Latency for measuring CX and Network Optimization" [Ghai2021]

- Robin Marx and Joris Herbots. "Merge Those Metrics: Towards Holistic (Protocol) Logging" [Marx2021]

- Sandor Laki, Szilveszter Nadas, Balazs Varga, and Luis M. Contreras. "Incentive-Based Traffic Management and QoS Measurements" [Laki2021]

- Satadal Sengupta, Hyojoon Kim, and Jennifer Rexford. "Fine-Grained RTT Monitoring Inside the Network" [Sengupta2021]

- Stuart Cheshire. "The Internet is a Shared Network" [Cheshire2021]

- Toerless Eckert and Alex Clemm. "network-quality-eckert-clemm-00.4"

- Vijay Sivaraman, Sharat Madanapalli, and Himal Kumar. "Measuring Network Experience Meaningfully, Accurately, and Scalably" [Sivaraman2021]

• Yaakov (J) Stein. "The Futility of QoS" [Stein2021]

4. Workshop Topics and Discussion

The agenda for the three-day workshop was broken into four separate sections that each played a role in framing the discussions. The workshop started with a series of introduction and problem space presentations (Section 4.1), followed by metrics considerations (Section 4.2), cross-layer considerations (Section 4.3), and a synthesis discussion (Section 4.4). After the four subsections concluded, a follow-on discussion was held to draw conclusions that could be agreed upon by workshop participants (Section 5).

4.1. Introduction and Overviews

The workshop started with a broad focus on the state of user Quality of Service (QoS) and Quality of Experience (QoE) on the Internet today. The goal of the introductory talks was to set the stage for the workshop by describing both the problem space and the current solutions in place and their limitations.

The introduction presentations provided views of existing QoS and QoE measurements and their effectiveness. Also discussed was the interaction between multiple users within the network, as well as the interaction between multiple layers of the OSI stack. Vint Cerf provided a keynote describing the history and importance of the topic.

4.1.1. Key Points from the Keynote by Vint Cerf

We may be operating in a networking space with dramatically different parameters compared to 30 years ago. This differentiation justifies reconsidering not only the importance of one metric over the other but also reconsidering the entire metaphor.

It is time for the experts to look at not only adjusting TCP but also exploring other protocols, such as QUIC has done lately. It's important that we feel free to consider alternatives to TCP. TCP is not a teddy bear, and one should not be afraid to replace it with a transport layer with better properties that better benefit its users.

A suggestion: we should consider exercises to identify desirable properties. As we are looking at the parametric spaces, one can identify "desirable properties", as opposed to "fundamental properties", for example, a low-latency property. An example coming from the Advanced Research Projects Agency (ARPA): you want to know where the missile is now, not where it was. Understanding drives particular parameter creation and selection in the design space.

When parameter values are changed in extreme, such as connectiveness, alternative designs will emerge. One case study of note is the interplanetary protocol, where "ping" is no longer indicative of anything useful. While we look at responsiveness, we should not ignore connectivity.

Unfortunately, maintaining backward compatibility is painful. The work on designing IPv6 so as to transition from IPv4 could have been done better if the backward compatibility was considered. It is too late for IPv6, but it is not too late to consider this issue for potential future problems.

IPv6 is still not implemented fully everywhere. It's been a long road to deployment since starting work in 1996, and we are still not there. In 1996, the thinking was that it was quite easy to implement IPv6, but that failed to hold true. In 1996, the dot-com boom began, where a lot of money was spent quickly, and the moment was not caught in time while the market expanded exponentially. This should serve as a cautionary tale.

One last point: consider performance across multiple hops in the Internet. We've not seen many end-to-end metrics, as successfully developing end-to-end measurements across different network and business boundaries is quite hard to achieve. A good question to ask when developing new protocols is "will the new protocol work across multiple network hops?"

Multi-hop networks are being gradually replaced by humongous, flat networks with sufficient connectivity between operators so that systems become 1 hop, or 2 hops at most, away from each other (e.g., Google, Facebook, and Amazon). The fundamental architecture of the Internet is changing.

4.1.2. Introductory Talks

The Internet is a shared network built on IP protocols using packet switching to interconnect multiple autonomous networks. The Internet's departure from circuit-switching technologies allowed it to scale beyond any other known network design. On the other hand, the lack of in-network regulation made it difficult to ensure the best experience for every user.

As Internet use cases continue to expand, it becomes increasingly more difficult to predict which network characteristics correlate with better user experiences. Different application classes, e.g., video streaming and teleconferencing, can affect user experience in ways that are complex and difficult to measure. Internet utilization shifts rapidly during the course of each day, week, and year, which further complicates identifying key metrics capable of predicting a good user experience.

QoS initiatives attempted to overcome these difficulties by strictly prioritizing different types of traffic. However, QoS metrics do not always correlate with user experience. The utility of the QoS metric is further limited by the difficulties in building solutions with the desired QoS characteristics.

QoE initiatives attempted to integrate the psychological aspects of how quality is perceived and create statistical models designed to optimize the user experience. Despite these high modeling efforts, the QoE approach proved beneficial in certain application classes. Unfortunately, generalizing the models proved to be difficult, and the question of how different applications affect each other when sharing the same network remains an open problem.

The industry's focus on giving the end user more throughput/bandwidth led to remarkable advances. In many places around the world, a home user enjoys gigabit speeds to their ISP. This is so remarkable that it would have been brushed off as science fiction a decade ago. However, the focus on increased capacity came at the expense of neglecting another important core metric: latency. As a result, end users whose experience is negatively affected by high latency were advised to upgrade their equipment to get more throughput instead. [MacMillian2021] showed that sometimes such an upgrade can lead to latency improvements, due to the economical reasons of overselling the "value-priced" data plans.

As the industry continued to give end users more throughput, while mostly neglecting latency concerns, application designs started to employ various latency and short service disruption hiding techniques. For example, a user's web browser performance experience is closely tied to the content in the browser's local cache. While such techniques can clearly improve the user experience when using stale data is possible, this development further decouples user experience from core metrics.

In the most recent 10 years, efforts by Dave Taht and the bufferbloat society have led to significant progress in updating queuing algorithms to reduce latencies under load compared to simpler FIFO queues. Unfortunately, the home router industry has yet to implement these algorithms, mostly due to marketing and cost concerns. Most home router manufacturers depend on System on a Chip (SoC) acceleration to create products with a desired throughput. SoC manufacturers opt for simpler algorithms and aggressive aggregation, reasoning that a higher-throughput chip will have guaranteed demand. Because consumers are offered choices primarily among different high-throughput devices, the perception that a higher throughput leads to higher a QoS continues to strengthen.

The home router is not the only place that can benefit from clearer indications of acceptable performance for users. Since users perceive the Internet via the lens of applications, it is important that we call upon application vendors to adopt solutions that stress lower latencies. Unfortunately, while bandwidth is straightforward to measure, responsiveness is trickier. Many applications have found a set of metrics that are helpful to their realm but do not generalize well and cannot become universally applicable. Furthermore, due to the highly competitive application space, vendors may have economic reasons to avoid sharing their most useful metrics.

4.1.3. Introductory Talks - Key Points

1. Measuring bandwidth is necessary but is not alone sufficient.
2. In many cases, Internet users don't need more bandwidth but rather need "better bandwidth", i.e., they need other connectivity improvements.
3. Users perceive the quality of their Internet connection based on the applications they use, which are affected by a combination of factors. There's little value in exposing a typical user to the entire spectrum of possible reasons for the poor performance perceived in their application-centric view.
4. Many factors affecting user experience are outside the users' sphere of control. It's unclear whether exposing users to these other factors will help them understand the state of their

network performance. In general, users prefer simple, categorical choices (e.g., "good", "better", and "best" options).

5. The Internet content market is highly competitive, and many applications develop their own "secret sauce".

4.2. Metrics Considerations

In the second agenda section, the workshop continued its discussion about metrics that can be used instead of or in addition to available bandwidth. Several workshop attendees presented deep-dive studies on measurement methodology.

4.2.1. Common Performance Metrics

Losing Internet access entirely is, of course, the worst user experience. Unfortunately, unless rebooting the home router restores connectivity, there is little a user can do other than contacting their service provider. Nevertheless, there is value in the systematic collection of availability metrics on the client side; these can help the user's ISP localize and resolve issues faster while enabling users to better choose between ISPs. One can measure availability directly by simply attempting connections from the client side to distant locations of interest. For example, Ookla's [Speedtest] uses a large number of Android devices to measure network and cellular availability around the globe. Ookla collects hundreds of millions of data points per day and uses these for accurate availability reporting. An alternative approach is to derive availability from the failure rates of other tests. For example, [FCC_MBA] and [FCC_MBA_methodology] use thousands of off-the-shelf routers, with measurement software developed by [SamKnows]. These routers perform an array of network tests and report availability based on whether test connections were successful or not.

Measuring available capacity can be helpful to end users, but it is even more valuable for service providers and application developers. High-definition video streaming requires significantly more capacity than any other type of traffic. At the time of the workshop, video traffic constituted 90% of overall Internet traffic and contributed to 95% of the revenues from monetization (via subscriptions, fees, or ads). As a result, video streaming services, such as Netflix, need to continuously cope with rapid changes in available capacity. The ability to measure available capacity in real time leverages the different adaptive bitrate (ABR) compression algorithms to ensure the best possible user experience. Measuring aggregated capacity demand allows ISPs to be ready for traffic spikes. For example, during the end-of-year holiday season, the global demand for capacity has been shown to be 5-7 times higher than during other seasons. For end users, knowledge of their capacity needs can help them select the best data plan given their intended usage. In many cases, however, end users have more than enough capacity, and adding more bandwidth will not improve their experience -- after a point, it is no longer the limiting factor in user experience. Finally, the ability to differentiate between the "throughput" and the "goodput" can be helpful in identifying when the network is saturated.

In measuring network quality, latency is defined as the time it takes a packet to traverse a network path from one end to the other. At the time of this report, users in many places worldwide can enjoy Internet access that has adequately high capacity and availability for their current needs. For these users, latency improvements, rather than bandwidth improvements,

can lead to the most significant improvements in QoE. The established latency metric is a round-trip time (RTT), commonly measured in milliseconds. However, users often find RTT values unintuitive since, unlike other performance metrics, high RTT values indicate poor latency and users typically understand higher scores to be better. To address this, [Paasch2021] and [Mathis2021] present an inverse metric, called "Round-trips Per Minute" (RPM).

There is an important distinction between "idle latency" and "latency under working conditions". The former is measured when the network is underused and reflects a best-case scenario. The latter is measured when the network is under a typical workload. Until recently, typical tools reported a network's idle latency, which can be misleading. For example, data presented at the workshop shows that idle latencies can be up to 25 times lower than the latency under typical working loads. Because of this, it is essential to make a clear distinction between the two when presenting latency to end users.

Data shows that rapid changes in capacity affect latency. [Foulkes2021] attempts to quantify how often a rapid change in capacity can cause network connectivity to become "unstable" (i.e., having high latency with very little throughput). Such changes in capacity can be caused by infrastructure failures but are much more often caused by in-network phenomena, like changing traffic engineering policies or rapid changes in cross-traffic.

Data presented at the workshop shows that 36% of measured lines have capacity metrics that vary by more than 10% throughout the day and across multiple days. These differences are caused by many variables, including local connectivity methods (Wi-Fi vs. Ethernet), competing LAN traffic, device load/configuration, time of day, and local loop/backhaul capacity. These factor variations make measuring capacity using only an end-user device or other end-network measurement difficult. A network router seeing aggregated traffic from multiple devices provides a better vantage point for capacity measurements. Such a test can account for the totality of local traffic and perform an independent capacity test. However, various factors might still limit the accuracy of such a test. Accurate capacity measurement requires multiple samples.

As users perceive the Internet through the lens of applications, it may be difficult to correlate changes in capacity and latency with the quality of the end-user experience. For example, web browsers rely on cached page versions to shorten page load times and mitigate connectivity losses. In addition, social networking applications often rely on prefetching their "feed" items. These techniques make the core in-network metrics less indicative of the users' experience and necessitates collecting data from the end-user applications themselves.

It is helpful to distinguish between applications that operate on a "fixed latency budget" from those that have more tolerance to latency variance. Cloud gaming serves as an example application that requires a "fixed latency budget", as a sudden latency spike can decide the "win/lose" ratio for a player. Companies that compete in the lucrative cloud gaming market make significant infrastructure investments, such as building entire data centers closer to their users. These data centers highlight the economic benefit that lower numbers of latency spikes outweigh the associated deployment costs. On the other hand, applications that are more tolerant to latency spikes can continue to operate reasonably well through short spikes. Yet, even those applications can benefit from consistently low latency depending on usage shifts. For example,

Video-on-Demand (VOD) apps can work reasonably well when the video is consumed linearly, but once the user tries to "switch a channel" or to "skip ahead", the user experience suffers unless the latency is sufficiently low.

Finally, as applications continue to evolve, in-application metrics are gaining in importance. For example, VOD applications can assess the QoE by application-specific metrics, such as whether the video player is able to use the highest possible resolution, identifying when the video is smooth or freezing, or other similar metrics. Application developers can then effectively use these metrics to prioritize future work. All popular video platforms (YouTube, Instagram, Netflix, and others) have developed frameworks to collect and analyze VOD metrics at scale. One example is the Scuba framework used by Meta [Scuba].

Unfortunately, in-application metrics can be challenging to use for comparative research purposes. First, different applications often use different metrics to measure the same phenomena. For example, application A may measure the smoothness of video via "mean time to rebuffer", while application B may rely on the "probability of rebuffering per second" for the same purpose. A different challenge with in-application metrics is that VOD is a significant source of revenue for companies, such as YouTube, Facebook, and Netflix, placing a proprietary incentive against exchanging the in-application data. A final concern centers on the privacy issues resulting from in-application metrics that accurately describe the activities and preferences of an individual end user.

4.2.2. Availability Metrics

Availability is simply defined as whether or not a packet can be sent and then received by its intended recipient. Availability is naively thought to be the simplest to measure, but it is more complex when considering that continual, instantaneous measurements would be needed to detect the smallest of outages. Also difficult is determining the root cause of infallibility: was the user's line down, was something in the middle of the network, or was it the service with which the user was attempting to communicate?

4.2.3. Capacity Metrics

If the network capacity does not meet user demands, the network quality will be impacted. Once the capacity meets the demands, increasing capacity won't lead to further quality improvements.

The actual network connection capacity is determined by the equipment and the lines along the network path, and it varies throughout the day and across multiple days. Studies involving DSL lines in North America indicate that over 30% of the DSL lines have capacity metrics that vary by more than 10% throughout the day and across multiple days.

Some factors that affect the actual capacity are:

1. Presence of a competing traffic, either in the LAN or in the WAN environments. In the LAN setting, the competing traffic reflects the multiple devices that share the Internet connection. In the WAN setting, the competing traffic often originates from the unrelated network flows that happen to share the same network path.

2. Capabilities of the equipment along the path of the network connection, including the data transfer rate and the amount of memory used for buffering.

3. Active traffic management measures, such as traffic shapers and policers that are often used by the network providers.

There are other factors that can negatively affect the actual line capacities.

The user demands of the traffic follow the usage patterns and preferences of the particular users. For example, large data transfers can use any available capacity, while the media streaming applications require limited capacity to function correctly. Videoconferencing applications typically need less capacity than high-definition video streaming.

4.2.4. Latency Metrics

End-to-end latency is the time that a particular packet takes to traverse the network path from the user to their destination and back. The end-to-end latency comprises several components:

1. The propagation delay, which reflects the path distance and the individual link technologies (e.g., fiber vs. satellite). The propagation doesn't depend on the utilization of the network, to the extent that the network path remains constant.

2. The buffering delay, which reflects the time segments spent in the memory of the network equipment that connect the individual network links, as well as in the memory of the transmitting endpoint. The buffering delay depends on the network utilization, as well as on the algorithms that govern the queued segments.

3. The transport protocol delays, which reflect the time spent in retransmission and reassembly, as well as the time spent when the transport is "head-of-line blocked".

4. Some of the workshop submissions that have explicitly called out the application delay, which reflects the inefficiencies in the application layer.

Typically, end-to-end latency is measured when the network is idle. Results of such measurements mostly reflect the propagation delay but not other kinds of delay. This report uses the term "idle latency" to refer to results achieved under idle network conditions.

Alternatively, if the latency is measured when the network is under its typical working conditions, the results reflect multiple types of delays. This report uses the term "working latency" to refer to such results. Other sources use the term "latency under load" (LUL) as a synonym.

Data presented at the workshop reveals a substantial difference between the idle latency and the working latency. Depending on the traffic direction and the technology type, the working latency is between 6 to 25 times higher than the idle latency:

Direction	Technology Type	Working Latency	Idle Latency	Working - Idle Difference	Working / Idle Ratio
Downstream	FTTH	148	10	138	15

Direction	Technology Type	Working Latency	Idle Latency	Working - Idle Difference	Working / Idle Ratio
Downstream	Cable	103	13	90	8
Downstream	DSL	194	10	184	19
Upstream	FTTH	207	12	195	17
Upstream	Cable	176	27	149	6
Upstream	DSL	686	27	659	25

Table 1

While historically the tooling available for measuring latency focused on measuring the idle latency, there is a trend in the industry to start measuring the working latency as well, e.g., Apple's [NetworkQuality].

4.2.5. Measurement Case Studies

The participants have proposed several concrete methodologies for measuring the network quality for the end users.

[Paasch2021] introduced a methodology for measuring working latency from the end-user vantage point. The suggested method incrementally adds network flows between the user device and a server endpoint until a bottleneck capacity is reached. From these measurements, a round-trip latency is measured and reported to the end user. The authors chose to report results with the RPM metric. The methodology had been implemented in Apple's macOS Monterey.

[Mathis2021] applied the RPM metric to the results of more than 4 billion download tests that M-Lab performed from 2010-2021. During this time frame, the M-Lab measurement platform underwent several upgrades that allowed the research team to compare the effect of different TCP congestion control algorithms (CCAs) on the measured end-to-end latency. The study showed that the use of cubic CCA leads to increased working latency, which is attributed to its use of larger queues.

[Schlinker2019] presented a large-scale study that aimed to establish a correlation between goodput and QoE on a large social network. The authors performed the measurements at multiple data centers from which video segments of set sizes were streamed to a large number of end users. The authors used the goodput and throughput metrics to determine whether particular paths were congested.

[Reed2021] presented the analysis of working latency measurements collected as part of the Measuring Broadband America (MBA) program by the Federal Communication Commission (FCC). The FCC does not include working latency in its yearly report but does offer it in the raw data files. The authors used a subset of the raw data to identify important differences in the working latencies across different ISPs.

[MacMillian2021] presented analysis of working latency across multiple service tiers. They found that, unsurprisingly, "premium" tier users experienced lower working latency compared to a "value" tier. The data demonstrated that working latency varies significantly within each tier; one possible explanation is the difference in equipment deployed in the homes.

These studies have stressed the importance of measurement of working latency. At the time of this report, many home router manufacturers rely on hardware-accelerated routing that uses FIFO queues. Focusing on measuring the working latency measurements on these devices and making the consumer aware of the effect of choosing one manufacturer vs. another can help improve the home router situation. The ideal test would be able to identify the working latency and pinpoint the source of the delay (home router, ISP, server side, or some network node in between).

Another source of high working latency comes from network routers exposed to cross-traffic. As [Schlinker2019] indicated, these can become saturated during the peak hours of the day. Systematic testing of the working latency in routers under load can help improve both our understanding of latency and the impact of deployed infrastructure.

4.2.6. Metrics Key Points

The metrics for network quality can be roughly grouped into the following:

1. Availability metrics, which indicate whether the user can access the network at all.
2. Capacity metrics, which indicate whether the actual line capacity is sufficient to meet the user's demands.
3. Latency metrics, which indicate if the user gets the data in a timely fashion.
4. Higher-order metrics, which include both the network metrics, such as inter-packet arrival time, and the application metrics, such as the mean time between rebuffering for video streaming.

The availability metrics can be seen as a derivative of either the capacity (zero capacity leading to zero availability) or the latency (infinite latency leading to zero availability).

Key points from the presentations and discussions included the following:

1. Availability and capacity are "hygienic factors" -- unless an application is capable of using extra capacity, end users will see little benefit from using over-provisioned lines.
2. Working latency has a stronger correlation with the user experience than latency under an idle network load. Working latency can exceed the idle latency by order of magnitude.
3. The RPM metric is a stable metric, with positive values being better, that may be more effective when communicating latency to end users.
4. The relationship between throughput and goodput can be effective in finding the saturation points, both in client-side [Paasch2021] and server-side [Schlinker2019] settings.
5. Working latency depends on the algorithm choice for addressing endpoint congestion control and router queuing.

Finally, it was commonly agreed to that the best metrics are those that are actionable.

4.3. Cross-Layer Considerations

In the cross-layer segment of the workshop, participants presented material on and discussed how to accurately measure exactly where problems occur. Discussion centered especially on the differences between physically wired and wireless connections and the difficulties of accurately determining problem spots when multiple different types of network segments are responsible for the quality. As an example, [Kerpez2021] showed that a limited bandwidth of 2.4 Ghz Wi-Fi bottlenecks the most frequently. In comparison, the wider bandwidth of the 5 Ghz Wi-Fi has only bottlenecked in 20% of observations.

The participants agreed that no single component of a network connection has all the data required to measure the effects of the network performance on the quality of the end-user experience.

- Applications that are running on the end-user devices have the best insight into their respective performance but have limited visibility into the behavior of the network itself and are unable to act based on their limited perspective.
- ISPs have good insight into QoS considerations but are not able to infer the effect of the QoS metrics on the quality of end-user experiences.
- Content providers have good insight into the aggregated behavior of the end users but lack the insight on what aspects of network performance are leading indicators of user behavior.

The workshop had identified the need for a standard and extensible way to exchange network performance characteristics. Such an exchange standard should address (at least) the following:

- A scalable way to capture the performance of multiple (potentially thousands of) endpoints.
- The data exchange format should prevent data manipulation so that the different participants won't be able to game the mechanisms.
- Preservation of end-user privacy. In particular, federated learning approaches should be preferred so that no centralized entity has the access to the whole picture.
- A transparent model for giving the different actors on a network connection an incentive to share the performance data they collect.
- An accompanying set of tools to analyze the data.

4.3.1. Separation of Concerns

Commonly, there's a tight coupling between collecting performance metrics, interpreting those metrics, and acting upon the interpretation. Unfortunately, such a model is not the best for successfully exchanging cross-layer data, as:

- actors that are able to collect particular performance metrics (e.g., the TCP RTT) do not necessarily have the context necessary for a meaningful interpretation,
- the actors that have the context and the computational/storage capacity to interpret metrics do not necessarily have the ability to control the behavior of the network/application, and

- the actors that can control the behavior of networks and/or applications typically do not have access to complete measurement data.

The participants agreed that it is important to separate the above three aspects, so that:

- the different actors that have the data, but not the ability to interpret and/or act upon it, should publish their measured data and
- the actors that have the expertise in interpreting and synthesizing performance data should publish the results of their interpretations.

4.3.2. Security and Privacy Considerations

Preserving the privacy of Internet end users is a difficult requirement to meet when addressing this problem space. There is an intrinsic trade-off between collecting more data about user activities and infringing on their privacy while doing so. Participants agreed that observability across multiple layers is necessary for an accurate measurement of the network quality, but doing so in a way that minimizes privacy leakage is an open question.

4.3.3. Metric Measurement Considerations

- The following TCP protocol metrics have been found to be effective and are available for passive measurement:
  - TCP connection latency measured using selective acknowledgment (SACK) or acknowledgment (ACK) timing, as well as the timing between TCP retransmission events, are good proxies for end-to-end RTT measurements.
  - On the Linux platform, the tcp_info structure is the de facto standard for an application to inspect the performance of kernel-space networking. However, there is no equivalent de facto standard for user-space networking.

- The QUIC and MASQUE protocols make passive performance measurements more challenging.
  - An approach that uses federated measurement/hierarchical aggregation may be more valuable for these protocols.
  - The QLOG format seems to be the most mature candidate for such an exchange.

4.3.4. Towards Improving Future Cross-Layer Observability

The ownership of the Internet is spread across multiple administrative domains, making measurement of end-to-end performance data difficult. Furthermore, the immense scale of the Internet makes aggregation and analysis of this difficult. [Marx2021] presented a simple logging format that could potentially be used to collect and aggregate data from different layers.

Another aspect of the cross-layer collaboration hampering measurement is that the majority of current algorithms do not explicitly provide performance data that can be used in cross-layer analysis. The IETF community could be more diligent in identifying each protocol's key performance indicators and exposing them as part of the protocol specification.

Despite all these challenges, it should still be possible to perform limited-scope studies in order to have a better understanding of how user quality is affected by the interaction of the different components that constitute the Internet. Furthermore, recent development of federated learning algorithms suggests that it might be possible to perform cross-layer performance measurements while preserving user privacy.

4.3.5. Efficient Collaboration between Hardware and Transport Protocols

With the advent of the low latency, low loss, and scalable throughput (L4S) congestion notification and control, there is an even higher need for the transport protocols and the underlying hardware to work in unison.

At the time of the workshop, the typical home router uses a single FIFO queue that is large enough to allow amortizing the lower-layer header overhead across multiple transport PDUs. These designs worked well with the cubic congestion control algorithm, yet the newer generation of algorithms can operate on much smaller queues. To fully support latencies less than 1 ms, the home router needs to work efficiently on sequential transmissions of just a few segments vs. being optimized for large packet bursts.

Another design trait common in home routers is the use of packet aggregation to further amortize the overhead added by the lower-layer headers. Specifically, multiple IP datagrams are combined into a single, large transfer frame. However, this aggregation can add up to 10 ms to the packet sojourn delay.

Following the famous "you can't improve what you don't measure" adage, it is important to expose these aggregation delays in a way that would allow identifying the source of the bottlenecks and making hardware more suitable for the next generation of transport protocols.

4.3.6. Cross-Layer Key Points

- Significant differences exist in the characteristics of metrics to be measured and the required optimizations needed in wireless vs. wired networks.
- Identification of an issue's root cause is hampered by the challenges in measuring multi-segment network paths.
- No single component of a network connection has all the data required to measure the effects of the complete network performance on the quality of the end-user experience.
- Actionable results require both proper collection and interpretation.
- Coordination among network providers is important to successfully improve the measurement of end-user experiences.
- Simultaneously providing accurate measurements while preserving end-user privacy is challenging.
- Passive measurements from protocol implementations may provide beneficial data.

4.4. Synthesis

Finally, in the synthesis section of the workshop, the presentations and discussions concentrated on the next steps likely needed to make forward progress. Of particular concern is how to bring forward measurements that can make sense to end users trying to select between various networking subscription options.

4.4.1. Measurement and Metrics Considerations

One important consideration is how decisions can be made and what actions can be taken based on collected metrics. Measurements must be integrated with applications in order to get true application views of congestion, as measurements over different infrastructure or via other applications may return incorrect results. Congestion itself can be a temporary problem, and mitigation strategies may need to be different depending on whether it is expected to be a short-term or long-term phenomenon. A significant challenge exists in measuring short-term problems, driving the need for continuous measurements to ensure critical moments and long-term trends are captured. For short-term problems, workshop participants debated whether an issue that goes away is indeed a problem or is a sign that a network is properly adapting and self-recovering.

Important consideration must be taken when constructing metrics in order to understand the results. Measurements can also be affected by individual packet characteristics -- differently sized packets typically have a linear relationship with their delay. With this in mind, measurements can be divided into a delay based on geographical distances, a packet-size serialization delay, and a variable (noise) delay. Each of these three sub-component delays can be different and individually measured across each segment in a multi-hop path. Variable delay can also be significantly impacted by external factors, such as bufferbloat, routing changes, network load sharing, and other local or remote changes in performance. Network measurements, especially load-specific tests, must also be run long enough to ensure that any problems associated with buffering, queuing, etc. are captured. Measurement technologies should also distinguish between upstream and downstream measurements, as well as measure the difference between end-to-end paths and sub-path measurements.

4.4.2. End-User Metrics Presentation

Determining end-user needs requires informative measurements and metrics. How do we provide the users with the service they need or want? Is it possible for users to even voice their desires effectively? Only high-level, simplistic answers like "reliability", "capacity", and "service bundling" are typical answers given in end-user surveys. Technical requirements that operators can consume, like "low-latency" and "congestion avoidance", are not terms known to and used by end users.

Example metrics useful to end users might include the number of users supported by a service and the number of applications or streams that a network can support. An example solution to combat networking issues include incentive-based traffic management strategies (e.g., an application requesting lower latency may also mean accepting lower bandwidth). User-perceived latency must be considered, not just network latency -- user experience in-application to in-

server latency and network-to-network measurements may only be studying the lowest-level latency. Thus, picking the right protocol to use in a measurement is critical in order to match user experience (for example, users do not transmit data over ICMP, even though it is a common measurement tool).

In-application measurements should consider how to measure different types of applications, such as video streaming, file sharing, multi-user gaming, and real-time voice communications. It may be that asking users for what trade-offs they are willing to accept would be a helpful approach: would they rather have a network with low latency or a network with higher bandwidth? Gamers may make different decisions than home office users or content producers, for example.

Furthermore, how can users make these trade-offs in a fair manner that does not impact other users? There is a tension between solutions in this space vs. the cost associated with solving these problems, as well as which customers are willing to front these improvement costs.

Challenges in providing higher-priority traffic to users centers around the ability for networks to be willing to listen to client requests for higher incentives, even though commercial interests may not flow to them without a cost incentive. Shared mediums in general are subject to oversubscribing, such that the number of users a network can support is either accurate on an underutilized network or may assume an average bandwidth or other usage metric that fails to be accurate during utilization spikes. Individual metrics are also affected by in-home devices from cheap routers to microwaves and by (multi-)user behaviors during tests. Thus, a single metric alone or a single reading without context may not be useful in assisting a user or operator to determine where the problem source actually is.

User comprehension of a network remains a challenging problem. Multiple workshop participants argued for a single number (potentially calculated with a weighted aggregation formula) or a small number of measurements per expected usage (e.g., a "gaming" score vs. a "content producer" score). Many agreed that some users may instead prefer to consume simplified or color-coded ratings (e.g., good/better/best, red/yellow/green, or bronze/gold/platinum).

4.4.3. Synthesis Key Points

- Some proposed metrics:
  - Round-trips Per Minute (RPM)
  - users per network
  - latency
  - 99% latency and bandwidth

- Median and mean measurements are distractions from the real problems.
- Shared network usage greatly affects quality.
- Long measurements are needed to capture all facets of potential network bottlenecks.
- Better-funded research in all these areas is needed for progress.
- End users will best understand a simplified score or ranking system.

5. Conclusions

During the final hour of the three-day workshop, statements that the group deemed to be summary statements were gathered. Later, any statements that were in contention were discarded (listed further below for completeness). For this document, the authors took the original list and divided it into rough categories, applied some suggested edits discussed on the mailing list, and further edited for clarity and to provide context.

5.1. General Statements

1. Bandwidth is necessary but not alone sufficient.

2. In many cases, Internet users don't need more bandwidth but rather need "better bandwidth", i.e., they need other improvements to their connectivity.

3. We need both active and passive measurements -- passive measurements can provide historical debugging.

4. We need passive measurements to be continuous, archivable, and queriable, including reliability/connectivity measurements.

5. A really meaningful metric for users is whether their application will work properly or fail because of a lack of a network with sufficient characteristics.

6. A useful metric for goodness must actually incentivize goodness -- good metrics should be actionable to help drive industries towards improvement.

7. A lower-latency Internet, however achieved, would benefit all end users.

5.2. Specific Statements about Detailed Protocols/Techniques

1. Round-trips Per Minute (RPM) is a useful, consumable metric.

2. We need a usable tool that fills the current gap between network reachability, latency, and speed tests.

3. End users that want to be involved in QoS decisions should be able to voice their needs and desires.

4. Applications are needed that can perform and report good quality measurements in order to identify insufficient points in network access.

5. Research done by regulators indicate that users/consumers prefer a simple metric per application, which frequently resolves to whether the application will work properly or not.

6. New measurements and QoS or QoE techniques should not rely only or depend on reading TCP headers.

7. It is clear from developers of interactive applications and from network operators that lower latency is a strong factor in user QoE. However, metrics are lacking to support this statement directly.

5.3. Problem Statements and Concerns

1. Latency mean and medians are distractions from better measurements.

2. It is frustrating to only measure network services without simultaneously improving those services.

3. Stakeholder incentives aren't aligned for easy wins in this space. Incentives are needed to motivate improvements in public network access. Measurements may be one step towards driving competitive market incentives.

4. For future-proof networking, it is important to measure the ecological impact of material and energy usage.

5. We do not have incontrovertible evidence that any one metric (e.g., latency or speed) is more important than others to persuade device vendors to concentrate on any one optimization.

5.4. No-Consensus-Reached Statements

Additional statements were discussed and recorded that did not have consensus of the group at the time, but they are listed here for completeness:

1. We do not have incontrovertible evidence that bufferbloat is a prevalent problem.

2. The measurement needs to support reporting localization in order to find problems. Specifically:

   ◦ Detecting a problem is not sufficient if you can't find the location.
   ◦ Need more than just English -- different localization concerns.

3. Stakeholder incentives aren't aligned for easy wins in this space.

6. Follow-On Work

There was discussion during the workshop about where future work should be performed. The group agreed that some work could be done more immediately within existing IETF working groups (e.g., IPPM, DetNet, and RAW), while other longer-term research may be needed in IRTF groups.

7. IANA Considerations

This document has no IANA actions.

8. Security Considerations

A few security-relevant topics were discussed at the workshop, including but not limited to:

- what prioritization techniques can work without invading the privacy of the communicating parties and
- how oversubscribed networks can essentially be viewed as a DDoS attack.

9. Informative References

[Aldabbagh2021]   Aldabbagh, A., "Regulatory perspective on measuring network quality for end-users", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/2021-09-07-Aldabbagh-Ofcom-presentiont-to-IAB-1v00-1.pdf>.

[Arkko2021]   Arkko, J. and M. Kühlewind, "Observability is needed to improve network quality", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/iab-position-paper-observability.pdf>.

[Balasubramanian2021]   Balasubramanian, P., "Transport Layer Statistics for Network Quality", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/transportstatsquality.pdf>.

[Briscoe2021]   Briscoe, B., White, G., Goel, V., and K. De Schepper, "A Single Common Metric to Characterize Varying Packet Delay", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/single-delay-metric-1.pdf>.

[Casas2021]   Casas, P., "10 Years of Internet-QoE Measurements Video, Cloud, Conferencing, Web and Apps. What do we need from the Network Side?", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/net_quality_internet_qoe_CASAS.pdf>.

[Cheshire2021]   Cheshire, S., "The Internet is a Shared Network", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/draft-cheshire-internet-is-shared-00b.pdf>.

[Davies2021]   Davies, N. and P. Thompson, "Measuring Network Impact on Application Outcomes Using Quality Attenuation", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/PNSol-et-al-Submission-to-Measuring-Network-Quality-for-End-Users-1.pdf>.

[DeSchepper2021]   De Schepper, K., Tilmans, O., and G. Dion, "Challenges and opportunities of hardware support for Low Queuing Latency without Packet Loss", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Nokia-IAB-Measuring-Network-Quality-Low-Latency-measurement-workshop-20210802.pdf>.

[Dion2021]   Dion, G., De Schepper, K., and O. Tilmans, "Focusing on latency, not throughput, to provide a better internet experience and network quality", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Nokia-IAB-Measuring-Network-Quality-Improving-and-focusing-on-latency-.pdf>.

[Fabini2021]   Fabini, J., "Network Quality from an End User Perspective", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Fabini-IAB-NetworkQuality.txt>.

[FCC_MBA]   FCC, "Measuring Broadband America", <https://www.fcc.gov/general/measuring-broadband-america>.

[FCC_MBA_methodology]   FCC, "Measuring Broadband America - Open Methodology", <https://www.fcc.gov/general/measuring-broadband-america-open-methodology>.

[Foulkes2021]   Foulkes, J., "Metrics helpful in assessing Internet Quality", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/IAB_Metrics_helpful_in_assessing_Internet_Quality.pdf>.

[Ghai2021]   Ghai, R., "Using TCP Connect Latency for measuring CX and Network Optimization", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/xfinity-wifi-ietf-iab-v2-1.pdf>.

[Iyengar2021]   Iyengar, J., "The Internet Exists In Its Use", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/The-Internet-Exists-In-Its-Use.pdf>.

[Kerpez2021]   Shafiei, J., Kerpez, K., Cioffi, J., Chow, P., and D. Bousaber, "Wi-Fi and Broadband Data", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Wi-Fi-Report-ASSIA.pdf>.

[Kilkki2021]   Kilkki, K. and B. Finley, "In Search of Lost QoS", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Kilkki-In-Search-of-Lost-QoS.pdf>.

[Laki2021]   Nadas, S., Varga, B., Contreras, L.M., and S. Laki, "Incentive-Based Traffic Management and QoS Measurements", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/11/CamRdy-IAB_user_meas_WS_Nadas_et_al_IncentiveBasedTMwQoS.pdf>.

[Liubogoshchev2021]   Liubogoshchev, M., "Cross-layer Cooperation for Better Network Service", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Cross-layer-Cooperation-for-Better-Network-Service-2.pdf>.

[MacMillian2021]   MacMillian, K. and N. Feamster, "Beyond Speed Test: Measuring Latency Under Load Across Different Speed Tiers", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/2021_nqw_lul.pdf>.

[Marx2021]   Marx, R. and J. Herbots, "Merge Those Metrics: Towards Holistic (Protocol) Logging", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/MergeThoseMetrics_Marx_Jul2021.pdf>.

[Mathis2021]   Mathis, M., "Preliminary Longitudinal Study of Internet Responsiveness", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Preliminary-Longitudinal-Study-of-Internet-Responsiveness-1.pdf>.

[McIntyre2021]   Paasch, C., McIntyre, K., Shapira, O., Meyer, R., and S. Cheshire, "An end-user approach to an Internet Score", September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Internet-Score-2.pdf>.

[Michel2021]   Michel, F. and O. Bonaventure, "Packet delivery time as a tie-breaker for assessing Wi-Fi access points", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/camera_ready_Packet_delivery_time_as_a_tie_breaker_for_assessing_Wi_Fi_access_points.pdf>.

[Mirsky2021]   Mirsky, G., Min, X., Mishra, G., and L. Han, "The error performance metric in a packet-switched network", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/IAB-worshop-Error-performance-measurement-in-packet-switched-networks.pdf>.

[Morton2021]   Morton, A. C., "Dream-Pipe or Pipe-Dream: What Do Users Want (and how can we assure it)?", Work in Progress, Internet-Draft, draft-morton-ippm-pipe-dream-01, 6 September 2021, <https://datatracker.ietf.org/doc/html/draft-morton-ippm-pipe-dream-01>.

[NetworkQuality]   Apple, "Network Quality", <https://support.apple.com/en-gb/HT212313>.

[Paasch2021]   Paasch, C., Meyer, R., Cheshire, S., and O. Shapira, "Responsiveness under Working Conditions", Work in Progress, Internet-Draft, draft-cpaasch-ippm-responsiveness-01, 25 October 2021, <https://datatracker.ietf.org/doc/html/draft-cpaasch-ippm-responsiveness-01>.

[Pardue2021]   Pardue, L. and S. Tellakula, "Lower-layer performance is not indicative of upper-layer success", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Lower-layer-performance-is-not-indicative-of-upper-layer-success-20210906-00-1.pdf>.

[Reed2021]   Reed, D.P. and L. Perigo, "Measuring ISP Performance in Broadband America: A Study of Latency Under Load", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Camera_Ready_-Measuring-ISP-Performance-in-Broadband-America.pdf>.

[SamKnows]   "SamKnows", <https://www.samknows.com/>.

[Schlinker2019]   Schlinker, B., Cunha, I., Chiu, Y., Sundaresan, S., and E. Katz-Basset, "Internet Performance from Facebook's Edge", February 2019, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Internet-Performance-from-Facebooks-Edge.pdf>.

[Scuba]   Abraham, L. et al., "Scuba: Diving into Data at Facebook", <https://research.facebook.com/publications/scuba-diving-into-data-at-facebook/>.

[Sengupta2021]   Sengupta, S., Kim, H., and J. Rexford, "Fine-Grained RTT Monitoring Inside the Network", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/Camera_Ready__Fine-Grained_RTT_Monitoring_Inside_the_Network.pdf>.

[Sivaraman2021]   Sivaraman, V., Madanapalli, S., and H. Kumar, "Measuring Network Experience Meaningfully, Accurately, and Scalably", February 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/CanopusPositionPaperCameraReady.pdf>.

[Speedtest]   Ookla, "Speedtest", <https://www.speedtest.net>.

[Stein2021]   Stein, Y., "The Futility of QoS", August 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/QoS-futility.pdf>.

[Welzl2021]   Welzl, M., "A Case for Long-Term Statistics", February 2021, <https://
www.iab.org/wp-content/IAB-uploads/2021/09/iab-
longtermstats_cameraready.docx-1.pdf>.

[WORKSHOP]   IAB, "IAB Workshop: Measuring Network Quality for End-Users, 2021",
September 2021, <https://www.iab.org/activities/workshops/network-quality>.

[Zhang2021]   Zhang, M., Goel, V., and L. Xu, "User-Perceived Latency to Measure CCAs",
September 2021, <https://www.iab.org/wp-content/IAB-uploads/2021/09/
User_Perceived_Latency-1.pdf>.

Appendix A.   Program Committee

The program committee consisted of:

> Jari Arkko
>
> Olivier Bonaventure
>
> Vint Cerf
>
> Stuart Cheshire
>
> Sam Crowford
>
> Nick Feamster
>
> Jim Gettys
>
> Toke Hoiland-Jorgensen
>
> Geoff Huston
>
> Cullen Jennings
>
> Katarzyna Kosek-Szott
>
> Mirja Kühlewind
>
> Jason Livingood
>
> Matt Mathis
>
> Randall Meyer
>
> Kathleen Nichols
>
> Christoph Paasch
>
> Tommy Pauly
>
> Greg White
>
> Keith Winstein

Appendix B.   Workshop Chairs

The workshop chairs consisted of:

> Wes Hardaker
>
> Evgeny Khorov
>
> Omer Shapira

Appendix C.   Workshop Participants

The following is a list of participants who attended the workshop over a remote connection:

> Ahmed Aldabbagh
>
> Jari Arkko
>
> Praveen Balasubramanian
>
> Olivier Bonaventure
>
> Djamel Bousaber
>
> Bob Briscoe
>
> Rich Brown
>
> Anna Brunstrom
>
> Pedro Casas
>
> Vint Cerf
>
> Stuart Cheshire
>
> Kenjiro Cho
>
> Steve Christianson
>
> John Cioffi
>
> Alexander Clemm
>
> Luis M. Contreras
>
> Sam Crawford
>
> Neil Davies
>
> Gino Dion
>
> Toerless Eckert
>
> Lars Eggert
>
> Joachim Fabini
>
> Gorry Fairhurst

Nick Feamster

Mat Ford

Jonathan Foulkes

Jim Gettys

Rajat Ghai

Vidhi Goel

Wes Hardaker

Joris Herbots

Geoff Huston

Toke Høiland-Jørgensen

Jana Iyengar

Cullen Jennings

Ken Kerpez

Evgeny Khorov

Kalevi Kilkki

Joon Kim

Zhenbin Li

Mikhail Liubogoshchev

Jason Livingood

Kyle MacMillan

Sharat Madanapalli

Vesna Manojlovic

Robin Marx

Matt Mathis

Jared Mauch

Kristen McIntyre

Randall Meyer

François Michel

Greg Mirsky

Cindy Morgan

Al Morton

Szilveszter Nadas

Kathleen Nichols

Lai Yi Ohlsen

Christoph Paasch

Lucas Pardue

Tommy Pauly

Levi Perigo

David Reed

Alvaro Retana

Roberto

Koen De Schepper

David Schinazi

Brandon Schlinker

Eve Schooler

Satadal Sengupta

Jinous Shafiei

Shapelez

Omer Shapira

Dan Siemon

Vijay Sivaraman

Karthik Sundaresan

Dave Taht

Rick Taylor

Bjørn Ivar Teigen

Nicolas Tessares

Peter Thompson

Balazs Varga

Bren Tully Walsh

Michael Welzl

Greg White

Russ White

Keith Winstein

Lisong Xu

Jiankang Yao

Gavin Young

Mingrui Zhang

IAB Members at the Time of Approval

Internet Architecture Board members at the time this document was approved for publication were:

Jari Arkko

Deborah Brungard

Lars Eggert

Wes Hardaker

Cullen Jennings

Mallory Knodel

Mirja Kühlewind

Zhenbin Li

Tommy Pauly

David Schinazi

Russ White

Qin Wu

Jiankang Yao

Acknowledgments

The authors would like to thank the workshop participants, the members of the IAB, and the program committee for creating and participating in many interesting discussions.

Contributors

Thank you to the people that contributed edits to this document:

Erik Auerswald

Simon Leinen

Brian Trammell

Authors' Addresses

Wes Hardaker
Email: ietf@hardakers.net

Omer Shapira
Email: omer_shapira@apple.com