Measuring Network Experience Meaningfully, Accurately, and Scalably

Vijay Sivaraman , Sharat Madanapalli , Himal Kumar
UNSW Sydney and Canopus Networks Pty Ltd, Australia

Abstract: We take the position that end-users should not be expected to understand network quality in terms of engineering metrics such as throughput, latency and loss. A typical user cares how well their network supports “experience” on their “application” - i.e. how often will their streaming video freeze, their game lag spike, and their conference stutter - and they need to be told this directly rather than by inference. We further contend that emerging capabilities in Programmable Networking and Machine Learning make it viable to measure application experience in a meaningful, accurate, and scalable manner. We briefly comment on recent academic research supporting our thesis, and on our experiences building and deploying a commercial platform to directly measure application experience at scale in carrier networks.

1. Application Experience Metrics Meaningful to Users

Network performance is reported today in terms of throughput - speed-test services like Ookla publish global indices [1] of average mobile and broadband speeds per country, and government regulators e.g. ACCC in Australia [2] operate nationally funded broadband speed-measurement programs. It is unclear if the outcomes are meaningful to a lay user; for example consider the headline figure below taken from the ACCC’s Measuring Broadband Australia (MBA) latest monthly report:

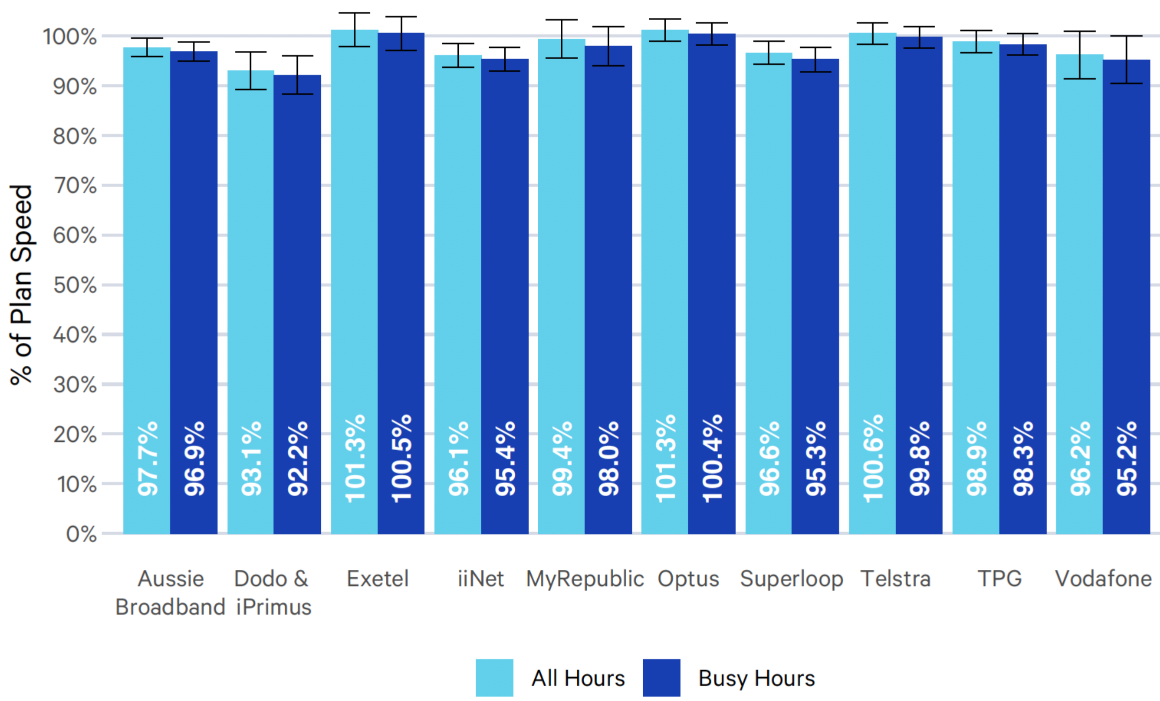

Fig. 1: Average download speed by ISP, taken from the ACCC MBA Report 14, Aug 2021 [2]

A lay user would infer from the figure above: “if I choose Telstra as my ISP, I will get 99.8% of my plan speed on average during peak hour, whereas if I choose Optus, I will get 100.4%”. A user would have no idea whether the reported speed difference across ISPs is significant or not to their streaming video, gaming, or conferencing experience. Add to this the variability of speed-test results (the error bars are larger than the differences across providers), their susceptibility to test conditions (such as server location, number of test threads, etc.), and their synthetic bursty traffic patterns that are unrepresentative of typical application behavior, the value of throughput measurement is indeed questionable [3].

As broadband speeds approach 100 Mbps, further increases in speed become largely imperceptible to users [4]. Indeed, growth in broadband speed is eclipsed by the growth in volume - the average Australian household consumed 355 GB in Dec 2020, representing a 59% year-on-year increase [5]. Our Internet usage is therefore increasingly resembling a marathon of video streaming, gaming, and teleconferencing consuming more and more data miles each day, rather than a few-second sprint as measured and reported using speed -tests.

In contrast to the speed-test comparison shown above, imagine if we could present the user with a direct comparison of streaming video and gaming experience across ISPs:

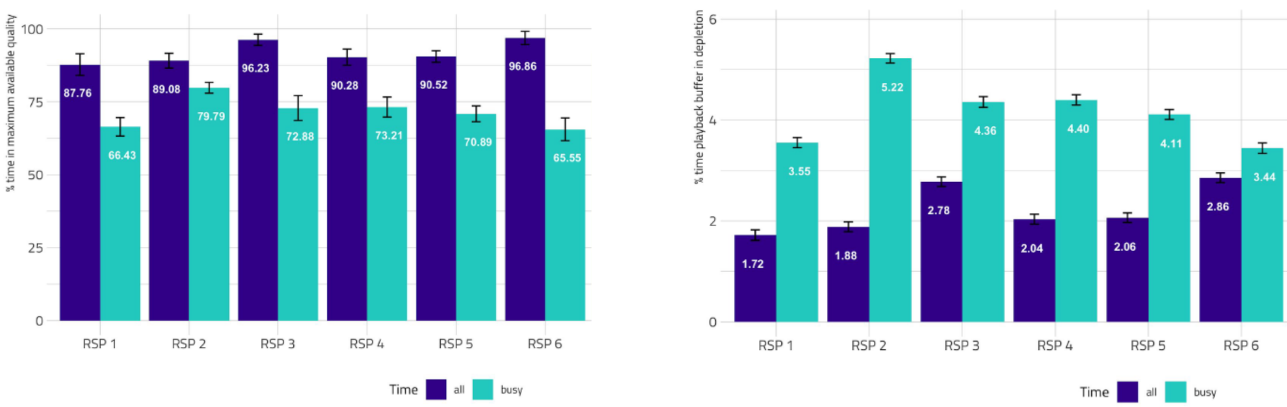

Fig. 2: Video streaming across ISPs: (a) % time in highest quality; and (b) buffer stalls per hour

Fig. 3: Lag spikes (90-th percentile latency) for selected game titles across ISPs

A lay person can easily glean from Fig. 2 that users are most likely to get the highest resolution video during busy hours from ISP2, while buffer stalls are lowest with ISP1. Fig. 3 informs them that ISP9 offers the most stable latency (i.e. lowest jitter) for online shooting games. Similar comparisons can be shown for conferencing apps [6], game downloads, etc.

Speed-test metrics can be enriched (if not replaced) with the application-level metrics above, since end-users can much more easily relate the latter with their everyday usage of the Internet.

2. Measuring Application Experience Accurately at Scale

Our second contention is that application experience meaningful to users (as described above) can be measured accurately and at scale, using emerging technologies of Programmable Networks and Machine Learning. Prior academic works such as [7,8,9] have shown that streaming video experience including stream resolution and buffer health can be estimated accurately by analysing network traffic, and recently in [10] we have shown that similar inferences can be made for live video streams. These methods work agnostic to packet encryption, using network behavioral patterns - video streams transfer data in chunks, with conspicuous spurts of network activity separated by idle periods - machine learning models can be trained to correlate the video resolution quality and buffer state with the time-series patterns of height, width, and spacing of chunks across potentially multiple flows corresponding to a video stream. Similar techniques have been developed for inferring online gaming experience [11] - state machines track the various stages (lobby, matchmaking, server-selection, and game-play) of each gaming stream, along with latency and jitter (using multiple techniques adapted to specific TCP and UDP games), and map them to experience values specific to the game genre (e.g shooting, strategy, sports, role-play, etc.). In (as yet) unpublished work we have extended our machines to deduce user experience for the major conferencing application platforms such as Zoom, Teams, WebEx, WhatsApp, and Discord.

Further, the application experience measurements above can be done at Terabit speeds using emerging programmable networks switches. Canopus Networks (https://www.canopusnet.com/) has built a commercial platform that: (a) takes a raw feed of Terabit traffic by optically tapping links in a carrier network; (b) extracts stream-level attributes (named FlowPulse^{TM}) using custom P4-code operating on a Intel/Barefoot Tofino-powered Programmable Switch; (c) exports these attributes via push (postcard-based) telemetry to ML engines in general purpose compute; (d) makes inferences on application type (video on-demand, live video, gaming, conferencing, downloads, etc.) and provider (Netflix, CoD, Zoom, Steam, etc.) for each long-lived traffic flow; and (e) estimates user experience in easy-to-understand terms (resolution and buffer health for streaming video, latency for gaming, stutters for conferencing, download speeds for gaming updates, etc.). These are exported via APIs for real-time consumption, and logged in databases for forensic evaluation.

3. Deployments and Learnings

The Canopus platform for application experience measurement is commercially deployed in multiple carrier networks in Australia, covering hundreds of thousands of broadband subscribers. Our ongoing engagements and preliminary results show promise in helping ISPs address various blind spots:

1. ***Better network dimensioning:*** Network operators in Australia need to purchase network capacity on a daily/weekly basis from the nationalised access infrastructure (NBN); data on usage behavior patterns coupled with quantification of user experience is allowing operators to minimise over-provisioning (which incurs unnecessary expense) while avoiding under-provisioning (which impacts user experience). The data is also helping them prepare for shock events, such as release of large game patches.
2. ***Quantifying premium gaming products:*** Network operators are starting to offer premium gaming products, such as gaming-optimised home routers and gaming VPN/CDN overlays. Data on user experience is helping operators benchmark the efficacy of their premium offerings in combating various factors affecting gameplay experience, including contention in the home, congestion in the aggregation network, and dynamic selection of game servers.
3. ***Identifying root cause of network problems:*** Customer support and churn affects the ISP bottom line. Visibility into user experience is enabling operators to trouble-shoot specific customer complaints relating to streaming, gaming, and conferencing experience degradation, by giving them fine-grained data on the potential root cause of the degradation event. In some instances it was found to be household contention (devices doing backups and cloud sync unbeknownst to the user), and in other instances it was determined to be more likely the home WiFi.
4. ***Network tuning:*** An interesting use-case enabled by the real-time experience measurement capability is to tune the network to achieve an explicit trade-off point in performance across applications. For example, tuning the shaping buffer in the BNG to a larger value improves the throughput of a download stream (including traffic blasts from a speed-test), but worsens jitter for a game-play stream (which sends periodically spaced packets). The ability to explicitly capture this application experience trade-off allows operators to make better network tuning decisions.

4. Conclusions

We would like to stimulate discussion to focus network quality measurement effort directly on user application experience, which we believe is meaningful to consumers, operators, and regulators. We are confident that this can be done accurately and at scale, using emerging Machine Learning and Programmable Networks technologies. A growing body of academic research is developing theoretical models for application experience inference agnostic to packet encryption, and we have built a commercial platform currently operational in multiple carrier networks. The data thus generated is helping inform several use-cases relating to network dimensioning, configuration tuning, customer support, and premium offerings, and we believe more value can be unlocked in the near future.

References

[1] Speedtest Global Index, https://www.speedtest.net/global-index
[2] Measuring Broadband Australia Program, https://www.accc.gov.au/regulated-infrastructure/communications/monitoring-reporting/measuring-broadband-australia-program
[3] Nick Feamster, Jason Livingood (2020). Measuring Internet Speed: Current Challenges and Future Recommendations. *Communications of the ACM (CACM), 63(12), pp. 72–80.*
[4] BBF Blog, "An economic argument for moving away from Mbps", https://www.broadband-forum.org/an-economic-argument-for-moving-away-from-mbps
[5] ACCC Internet Activity Report, Jun 2021, https://www.accc.gov.au/system/files/Internet%20Activity%20Report%20%28December%202020%29.pdf
[6] MacMillan, K., Mangla, T., Saxon, J., & Feamster, N. (2021). Measuring the Performance and Network Utilization of Popular Video Conferencing Applications. *arXiv:2105.13478*.
[7] T. Mangla, E. Halepovic, M. Ammar, E. Zegura, "eMIMIC: Estimating HTTP-Based Video QoE Metrics from Encrypted Network Traffic", IFIP Traffic Measurement and Analysis (TMA) Conference, Vienna, Austria, Jun 2018.
[8] S. Madanapalli, H. Habibi Gharakheili and V. Sivaraman, "Inferring Netflix User Experience from Broadband Network Measurement", IFIP Traffic Measurement and Analysis (TMA) Conference, Paris, France, Jun 2019.
[9] P. Schmitt, F. Bronzino, S. Ayoubi, G. Martins, R. Teixeira, N. Feamster, "Inferring Streaming Video Quality from Encrypted Traffic: Practical Models and Deployment Experience", ACM SIGMETRICS, pp. 1–25, Boston, Massachusetts, USA, Jun 2020.
[10] S. Madanapalli, A. Mathai, H. Habibi Gharakheili, V. Sivaraman, "ReCLive: Real-Time Classification and QoE Inference of Live Video Streaming Services", IEEE IWQoS, Tokyo, Japan, Jun 2021.
[11] V. Sivaraman and S. Madanapalli, "Network Telemetry for Measuring and Enhancing Online Gaming Experience", Australian Network Operators Group (AusNOG), Sep 2019, https://www.ausnog.net/sites/default/files/ausnog-2019/presentations/2.1_Vijay_Sivaraman_AusNOG2019.pdf