

QoE: A Micro and Macro Perspective

Henning Schulzrinne, Kyung-Hwa Kim, Hyunwoo Nam, Vishal Singh, Daniel Song
Columbia University, New York, NY

September 13, 2015

Abstract

Quality of experience (QoE) takes a user-focused view of the Internet experience, derived from lower-layer metrics like quality of service, reliability and usability, measuring "the difference between what a customer expected, and what they actually received." (FY/c)

QoE can be usefully divided into a micro and macro aspect, covering the individualized and often short-term experience of an individual Internet user, and the aggregate QoE across a service territory, country or region. Just as macroeconomics builds on the aggregate behavior of households and firms, macro-QoE can usefully build on micro-QoE.

We believe that QoE-related research must be both descriptive and actionable, that is, allow users to determine the source of QoE problems, identify the responsible part of the network and, where possible, take corrective action. Often, for both residential and enterprise users, the first question to be answered is: "Is it just me?" At the macro level, the question turns to the diagnosis of wholesale infrastructure disruption, policy and economic failures or successes.

We present two examples that address both the "micro" and "macro" aspects. Our DYSWIS (Do You See What I See) prototype allows end users to detect and isolate QoE impairment, while our YouSlow [1] project uses application behavior to diagnose local problems and to establish the correlation between low-level network metrics and user behavior.

1 Introduction

In the context of the Internet, Quality of Experience (QoE) can be taken to include all aspects of the service experience. "Quality of Experience (QoE, or simply QX) measures the difference between what a customer expected, and what they actually received. Consider QoS as measuring service quality 'from the inside out' and QX as measuring quality 'from the outside in.'"¹ This would include facets such as interaction with customer support, installation, billing, perception of security, price-value perception and speed of restoration after service outages, along with fast-loading web pages, buffering-free cat videos and HD-quality voice calls. Metrics such as the American Customer Satisfaction Index (ACSI)² measure the total perception of various services. (For example, in June 2015, ACSI wrote: "Customer satisfaction with information services including subscription TV, Internet, wireless and fixed line telephone, and computer software drops 3.4 percent to an ACSI score of 68.8 on a 0 to 100 scale, the lowest level in seven years. According to the American Customer Satisfaction Index, customer satisfaction with subscription TV service deteriorates further, tying Internet service providers (ISPs) at 63 the worst score among 43 industries covered by the Index."³)

However, for this discussion, we will focus on a more narrow part of quality of experience, namely the performance and reliability of Internet applications as experienced by the (human) user⁴. QoE can usefully be divided into a micro and macro aspect, in analogy to micro and macro-economics. Just like microeconomics studies the behavior of individuals and firms, micro-QoE focuses on an individual application or possibly a household, as the unit of interaction with the service provider. In contrast, just like macroeconomics is the "study of economy-wide phenomena" [3], at the regional, national or global level, we consider macro-QoE as the

¹<https://www.findyourcloud.com/Articles/Quality-of-Experience/The-Difference-Between-Quality-of-Service-and-Quality-of-Experience/>

²<http://www.theacsi.org/>

³<http://www.theacsi.org/news-and-resources/press-releases/press-2015/press-release-telecommunications-and-information-2015>

⁴We note that relatively little attention has been directed at QoE as experienced by non-human consumers of Internet content, e.g., in IoT or speech and image recognition applications. See [2] for an early example of such an effort.

aggregate evaluation of how well the Internet serves user needs across a provider, nation or group of nations. Similar to the economic realm, micro-QoE and macro-QoE are related, but have separate concerns, models and goals. (“Microeconomics and macroeconomics are closely intertwined. Because changes in the overall economy arise from the decisions of millions of individuals, it is impossible to understand macroeconomic developments without considering the associated microeconomic decisions.” [3])

For micro-QoE, we might ask question such as: How does a user rate the experience of phone calls, entertainment video or applications? What is cause of a short or longer-term disruption or degradation in the usability of the service? Are other users experiencing the same problem or is the problem local to my device or household? How reliable is the service? How often and for how long is service interrupted or impaired? How long does it take to restore service after a natural disaster?

Unlike in the early days of the Internet, where “surfing the Internet” was a novel activity, consumers now likely expect essentially a transparent service, at least for in-home use, i.e., where the Internet “plumbing” is no more visible than the user has to think about the water treatment plant, water heater and sewer when taking a shower (California excepted). Thus, the goal is less likely to be a mean opinion score (MOS) rating but rather the presence of any noticeable impairments or temporary loss of service.

For macro-QoE, the focus shifts to larger aggregates, both in time and space: How does QoE compare between service providers, intra-country regions or countries? Does QoE, both average and variance, depend on time of day? How does QoE affect adoption of new applications, e.g., by ISP or over time? What impact does QoE have on application innovation? How do changes in end system algorithm or network features (e.g., SDN or priority services) affect QoE?

While the questions differ, designing tools that can address *both* types of questions can be helpful. For example, many tools whose authors care primarily about macro questions may incentivize adoption by providing users with the ability to measure and diagnose their own micro QoE. One can argue that successful commercial efforts at macro performance estimation rely on exactly that incentive, such as Netalyzr [4] or various network speedtest applications. Applications that provide only limited immediate benefit to users seem to struggle getting adopted beyond the “geek” demographic.

Below, we briefly describe two Columbia University projects that illustrate both micro and macro aspects of QoE and their intersection.

2 YouSlow: Observing the Impact of Video QoE in HTML5

Today’s popular video streaming services such as Netflix, Hulu and YouTube stream video content to viewers over HTTP or HTTPS. To provide smooth streaming, they use adaptive bitrate (ABR) streaming technologies where a video player dynamically adjusts video bitrates based on estimated network conditions, buffer occupancy and hardware specifications of viewers’ devices. Therefore, user-perceived video quality depends on how well the player selects the most appropriate bitrate. A viewer may experience frequent rebufferings (i.e., a video is paused and resumes playing repeatedly) when the player requests a higher bitrate than what can be supported by the network, or the viewer may be consigned to a low bitrate if the network capacity is underestimated. Hence, from OTT video service provider’s viewpoint, improving ABR heuristics is a key factor to enhancing video QoE.

To improve ABR streaming, it is important to analyze the impact of changes of ABR heuristics on video QoE. While traditional quality of service (QoS) based metrics (e.g., measuring TCP throughput, packet delay and jitter) can be used to pinpoint network impairments, the metrics may not accurately reflect the viewer’s watching experience. Thus, we believe that QoE monitoring systems should focus on application-layer instead of transport-layer events. We suggest monitoring live playback events directly from within video players rather than the network middle-boxes or routers. As a proof of concept, we have developed a lightweight Chrome web browser plug-in that monitors various playback events including start-up latency, rebufferings and bitrate changes directly from ABR players while viewers watch videos on the YouTube web site. So far, we have collected over 400,000 YouTube views from more than 900 viewers located in more than 100 countries.

We evaluate various QoE metrics by analyzing video abandonment rates in YouTube. An abandonment occurs when a viewer closes video in the middle of a playback due to lack of interest or the annoyance caused by playback events such as long start-up latency, frequent rebufferings and bitrate changes. We show that tracking rebuffering ratio and bitrate changes as a function of playback time is useful to quantify abandonment rates for short videos. Our measurements show that rebuffering events increase abandonment rate six times more than start-up latency; start-up latency is mostly caused by pre-roll ads in YouTube. Most interestingly, our

analysis shows that even increasing a bitrate during playback can annoy viewers; when the bitrate changes, they abandon videos more than four times as often. Further, we show that a single rebuffering event can cause abandonment rate three times higher than a single bitrate change.

3 DYSWIS: Collaborative Network Diagnostics

While operating systems and computers have generally become more user-friendly and reliable, Internet usage can still be frustrating – applications fail silently, things that worked yesterday do not work today, and failures are often transient. Indeed, Internet QoE is often circumscribed by these failures.

Compared to a few years ago, consumer Internet usage has changed in at least four aspects: Users now expect to connect to a wide variety of networks, from home and office networks to Wi-Fi hotspots and cellular networks. Applications have become more demanding, as almost every application, from online calendars to games, relies on Internet connectivity and a number of applications, such as Voice over Internet Protocol (VoIP) and video on demand (VoD), require consistent performance. Finally, such applications often rely on the proper functioning of up to half a dozen parties, from the local wireless network to DNS servers, content delivery networks (CDNs), and various middleboxes. For all of these components, professional assistance is either unavailable or expensive, with the result that most users need to become unwilling network administrators (or rely on their technically-savvy children or friends for assistance). Thus, users have no good way to know whom to call or what to try when things go wrong.

Most applications provide, at best, minimal support to help pinpoint potential sources of trouble. For example, if Web access is slow, the cause could be high packet loss on the local wireless network due to interference, an overloaded residential Internet connection, wide-area network problems, a misconfiguration in the Network Address Translation (NAT) box, or a remote server problem. The appropriate action differs in each case, ranging from using a third-party DNS server to simply waiting and hoping that the server recovers.

Motivated by these real-world problems, we have developed DYSWIS (“Do You See What I See?”) [5], a system for end users and enterprises to diagnose a range of network-related problems. DYSWIS differs from other approaches in relying on the assistance of other network users, modeling the common pattern where one person asks somebody close by, “Hey, is your Internet working?” Also, reflecting the proliferation of services, both standardized and proprietary, DYSWIS is designed to be extensible by users and third parties, such as vendors of applications. New probes and rule sets can be added to the running system.

4 Future Work

We envision a combination of the two systems to correlate user experience and network behavior. Thus, QoE impairments may trigger collaborative diagnostics to track the root cause.

References

- [1] Hyunwoo Nam, Kyung-Hwa Kim, Doru Calin, and Henning Schulzrinne. YouSlow: A performance analysis tool for adaptive bitrate video streaming. In *Proc. of SIGCOMM 2014 Posters and Demos*, Chicago, Illinois, August 2014.
- [2] Wenyu Jiang and Henning Schulzrinne. Speech recognition performance as an effective perceived quality predictor. In *IWQoS*, Miami Beach, May 2002.
- [3] N. Gregory Mankiw. *Principles of Economics*. Cengage Learning, 2014.
- [4] Christian Kreibich, Nicholas Weaver, Boris Nechaev, and Vern Paxson. Netalyzr: Illuminating the edge network. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC ’10, pages 246–259, New York, NY, USA, 2010. ACM.
- [5] Kyung Hwa Kim, Hyunwoo Nam, Vishal Singh, Daniel Song, and Henning Schulzrinne. DYSWIS: Crowdsourcing a home network diagnosis. In *Proc. of 23rd International Conference on Computer Communications and Networks (ICCCN 2104)*, Shanghai, China, August 2014.