A reactive crowdsourcing-based QoE monitoring platform – Term 2

Ricky K. P. Mok and kc claffy

1 Introduction

In the first term, we implemented the QUINCE platform, and successfully collected a large amount of QoE data. Building on this success, in the second term, we propose to continue and expand the scale of deployment. Furthermore, we will enrich the functionality of QUINCE to support more types of measurement and automate the assignment of measurement parameters.

2 Task 1: Expand the scale of measurement

In the second term, we aim to collect more than 10,000 QoE measurements. To support measurement at this scale, we will develop and apply enhancements in two major directions: 1) recruit and retain more subjects, 2) automate the process of selecting video materials and composing video playback scenarios.

Including more gamification elements. In the previous two crowdsourcing-based experiments, we successfully achieved fairly strong user engagement. We will continue to polish the QUINCE user interface to ensure subjects can smoothly execute the assigned measurement tasks. We will further enhance QUINCE by introducing two more gamification elements: badges and leaderboards. Badges provide subjects some intermediate goals before achieving their maximum reward limit and also increase a subjects sense of achievement. We will design the criteria for receiving badges and update the evolve of scores. We will also implement a leaderboard, which can stimulate competition among subjects and yield additional measurements.

Automating selection of video materials. Currently, for the simulated video streaming QoE test, we manually download from YouTube 5 High Definition (HD) video clips distributed using a Common Creative license, and re-encode them into DASH format. We also composed 4 simulation scenarios based on our prior experience in QoE measurement. For the YouTube video streaming test, we manually selected another 8 YouTube video clips to embed into QUINCE.

Subjects may grow bored after repeatedly evaluating QoE on the same set of video clips. We also received email feedback from a subject who expressed this feeling in the November trial. In the second term, we will investigate the use of the YouTube Data API (or similar API provided by other video streaming providers) to automate the video selection process. We will use predefined keywords or video categories to retrieve a list of videos, such as Sports and Music. We will set various parameters to narrow search results to return only high quality and popular videos.

When a certain video and scenario pair has collected sufficient samples (e.g., ¿ 30), we will crawl a new list of videos and download their highest-definition version and format them in CAIDAs server as DASH format. For the YouTube QoE assessment, we can simply update the list in our database.

Another challenge is to compose the simulation scenarios. In addition to relying on existing literature to manually design more scenarios, we will use the YouTube measurement results collected from CAIDAs Ark as a baseline to compose different scenarios. During network congestion events, we can observe significant performance degradation. We can use the streaming performance metrics to reproduce these worst case performance. We believe that this method can obtain a more realistic evaluation of video streaming QoE in todays Internet.

Expanding the coverage of video streaming platforms. In the first term, we have successfully assessed YouTube QoE by embedding video clips on QUINCE. We have found at least two more major video streaming platforms (Vimeo and Dailymotion) that provide video player APIs similar to the YouTube one that enables us to embed videos into QUINCE. Even though there are slight differences across APIs, we will instrument these APIs to expand the coverage of QoE assessments to these two video streaming platforms.

3   Task 2: Correlating user-reported QoE with NTT laboratory-based assessment and network data

In this task, we will leverage the synergy of this collaboration between CAIDA and NTT to investigate various factors influencing the QoE. In particular, we will investigate environmental and behavioral factors. Environmental factors include the subjects screen resolution/sizes, OS, and browser, which can affect the presentation of the video playback on a subjects computer. For example, the artifacts in low resolution video can be more prominent in large-size monitors than smaller ones. Behavioral factors are related to a subjects past experience, emotion, and perception. We collected some information using an online survey asking subjects about their video watching habits and tolerance to different impairments. In the second term, we will analyze the data and correlate these factors with QoE.

We will also cross-correlate with the laboratory-based data collected by NTT. We can sample a few scenarios that are not feasible via QUINCE measurement, e.g., narrow ranges of MOS scores, and reproduce them in the NTT lab. Because QUINCE records all events that happen during the video playback, we can precisely simulate the video streaming experience perceived by QUINCE subjects in the lab. As laboratory-based assessment is conducted in a controlled environment, we can isolate confounding factors and reduce the variance in the QoE ratings.

We will investigate the relationship between QoE degradation and congestion status at observed interconnects. We will ask subjects to perform traceroute to measure their forward path to video caches. We can use the traceroute data to infer the interdomain link traversed by the subject. By combining historical measurement data for the link as observed from CAIDAs Ark platform, we can investigate whether streaming performance was possibly influenced by interdomain congestion events. We will also adaptively adjust the cooling off period of the YouTube QoE measurement, in order to collect more data when congestion is most likely to occur.

4   Project timeline

We expect the project starts on March 1, 2019. We will spend around 2 months to complete each task. After accomplishing each goal, we will use 1 month to prepare and conduct experiments.

We anticipate to perform three experiment trials in the next term. We will conduct the first two experiments in May and September, 2019, respectively. The last trial will be a joint experiment with NTT as validation, and will be conducted on January and February, 2020.

Experiment trial I. The objective of this trial is to examine the engagement level of users. We expect each user will contribute more measurement data with a lower cost. For video streaming QoE measurement, we will compile a new set of simulation scenarios to examine the impacts of various level of impairments on the QoE. We will extend the collection video streaming performance and the QoE two more video streaming providers (Vimeo and Dailymotion). We anticipate to collect more than 5,000 QoE data in this trial.

Experiment trial II. We will test the automate video selection mechanism, which can increase the diversity of the genre of video clips. We will conduct QoE experiments on more than 100 different YouTube/Viemo/Dailymotion videos. The data can support us to investigate the correlation between the video content and the QoE. We will collect another 5,000 QoE data in this trial.

Joint experiment with NTT. We will examine the simulation scenarios composed using historical network/video streaming measurement data collected from Ark, and study the level of QoE degradation during network congestion. We will assist NTT to deploy QUINCE in the NTT lab and conduct experiment. The data can support us to compare the QoE obtained by crowdsourcing approach and the traditional laboratory-based approach. We will collect around 5,000 QoE data in this experiment.
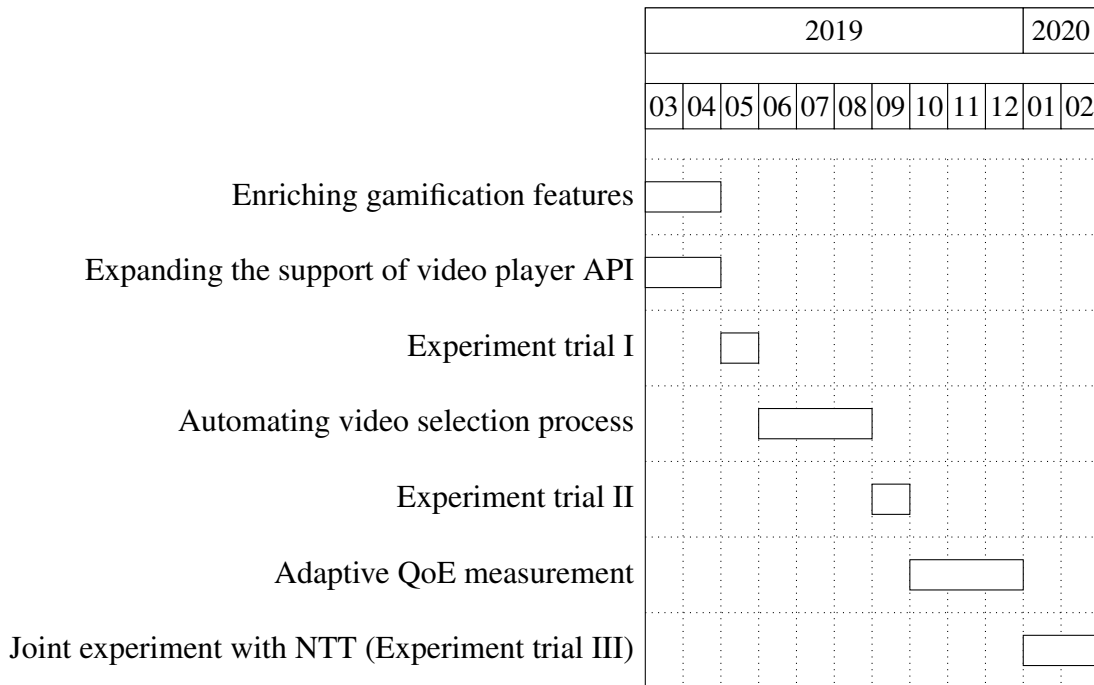

Figure 1: The Gantt chart illustrates the proposed schedule.