YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks

Pedro Casas
FTW Vienna
casas@ftw.at

Michael Seufert
FTW Vienna
seufert@ftw.at

Raimund Schatz
FTW Vienna
schatz@ftw.at

ABSTRACT

YouTube is changing the way operators manage network performance monitoring. In this paper we introduce YOUQ-MON, a novel on-line monitoring system for assessing the Quality of Experience (QoE) undergone by HSPA/3G customers watching YouTube videos, using network-layer measurements only. YOUQMON combines passive traffic analysis techniques to detect stalling events in YouTube video streams, with a QoE model to map stallings into a Mean Opinion Score reflecting the end-user experience. We evaluate the stalling detection performance of YOUQMON with hundreds of YouTube video streams, and present results showing the feasibility of performing real-time YouTube QoE monitoring in an operational mobile broadband network.

Keywords

QoE Monitoring, MOS, Stallings, YouTube, 3G Networks

1. INTRODUCTION

YouTube is the most popular application in today's Internet, accounting for more than 30% of the overall traffic worldwide. Its outstanding success poses a serious challenge for network operators, who need to engineer their systems to handle the huge volume of traffic in efficient ways. The issue becomes a real problem for mobile network operators, who need to offer high quality levels to reduce the risk of clients churning for quality dissatisfaction.

In this paper we present YOUQMON, a system capable of estimating the quality experienced by users watching YouTube videos from network traffic measurements only. YouTube QoE monitoring is about measuring the quality impairments perceived by the user at the application layer. From previous studies [1, 2] we know that the dominant QoE influence impairments of an HTTP video streaming service like YouTube are the number and the duration of the playback *stallings*, i.e., the events when the player stops the playback. Stallings occur due to the depletion of the player's buffer at the user's terminal. Some approaches to detect YouTube stallings rely on browser plugins installed at the user's terminal [3]. However, this approach is complex, not scalable, and relies on the cooperation of the end-user. Hence, the big challenge and interest for a network operator willing to monitor the YouTube QoE of his customers is to detect such stalling patterns exclusively from the network
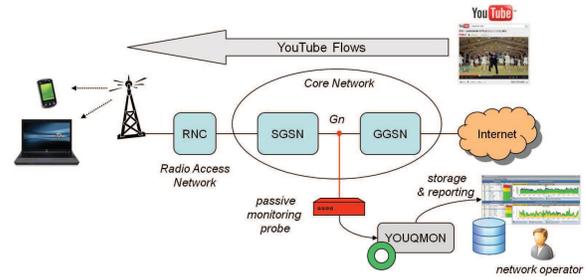
Figure 1: YOUQMON in a 3G Network.

side, being completely transparent to the end-users. This is exactly the approach followed by YOUQMON.

2. RELATED WORK

QoE monitoring in YouTube has very recently attracted the attention of the research community. Staehle et al. [3] introduced a client-side software tool to monitor YouTube traffic at the application layer, estimating the size of the YouTube playing buffer to predict stalling events on the video playback. Eckert et al. [4] presented an approach for measuring YouTube stallings in an off-line fashion, relying on network packet traces; the study presents some first interesting results, but the validation of the technique is limited to a couple of video examples. Schatz et al. [5] have recently introduced a monitoring technique to estimate the number of stalling events and their duration in a YouTube video, and tested its accuracy with off-line YouTube video packet traces. In this paper we build upon these studies and give a further step on the direction of QoE monitoring and assessment, presenting the first results on the large-scale, real-time monitoring of YouTube QoE in 3G/HSPA mobile broadband networks.

3. YOUQMON DESCRIPTION

YOUQMON works with packet data, passively captured at the vantage point of interest. Fig. 1 shows current experimental deployment of YOUQMON in the network of a major European mobile operator, using the METAWIN passive monitoring system [6] for data capturing, filtering, and analysis. Packets are captured on the Gn interface links between the Gateway GPRS Support Node (GGSN) and the Serving GPRS Support Node (SGSN). Intuitively, when the downlink bandwidth (DLW) or the throughput achieved by the YouTube flows (Th_{max}) are lower than the

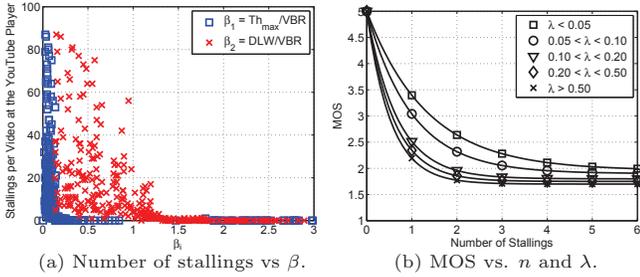(a) Number of stallings vs β.   (b) MOS vs. n and λ.

Figure 2: Stallings and QoE in YouTube.

corresponding video bitrate (VBR), the player buffer becomes gradually empty, ultimately leading to the stalling of the playback. However, measuring DLW and/or Th_{max} is not enough to get a precise picture of the user experience. Figure 2(a) shows the number of stallings measured at the YouTube player and it's correlation with the ratios $\beta_1 = Th_{max}/\text{VBR}$ and $\beta_2 = \text{DLW}/\text{VBR}$, for hundreds of different YouTube videos streamed to a local host with traffic shaping capabilities. While some marked relations can be observed for the average set of videos (e.g., no stallings if $\beta_1 > 0.3$ or $\beta_2 > 1.4$), it is difficult to judge the number of stallings for each specific video stream. For this reason, YOUQMON's QoE estimation principle is based on reconstructing the complete stalling pattern (i.e., number and duration of stallings) of a YouTube video from the analysis of the headers in the corresponding video packets.

YOUQMON is highly optimized to run on-line in standard 3G networks, as is its capable of decoding the complete 3GPP protocol stack in real-time. Nevertheless, the techniques apply to any IP-traffic vantage point. YOUQMON is fully developed in C for improved performance, and supports the two most popular video formats currently used by YouTube, namely Adobe-Flash and MPEG4.

The traffic analysis consists of two steps: (i) identify the beginning of every new YouTube video flow by HTTP header inspection, and (ii) extract the playtime offsets of the corresponding video frames to estimate the buffered video playtime at the YouTube player. The latter is achieved by inspecting the headers and tags of the video container (e.g., FLV or MP4). To preserve user privacy, any user related data (e.g., IMSI, MSISDN) are removed on-the-fly, and payload content is ignored. Using the extracted frame time offsets, YOUQMON tracks the amount of video playtime τ so far downloaded. The difference between τ and the current time of the video t corresponds to the remaining buffered video playtime $\Delta = \tau - t$. The video time t is computed as the difference between the actual time and the time t_p when the player starts to play the video; t_p is the time when the first packet containing actual video content is observed, plus an initial buffering time α_{play}. The video playback starts when the buffered playtime $\Delta > \alpha_{play}$, and stalls when $\Delta < \alpha_{sta}$. Both α_{play} and α_{sta} are estimated from large measurement campaigns, similar to [5]. By tracking the evolution of Δ, YOUQMON can estimate on the fly when the YouTube video is playing or stalled. Every 60 seconds, YOUQMON computes a new ticket containing the number of stallings n and the fraction of stalling time λ for each detected video. The final step consists in mapping the resulting stalling pattern into a Mean Opin-

ion Score (MOS) value. Using both lab and field subjective studies on YouTube QoE [1, 2], we have conceived a model that maps n and λ into a user experience measure: $\text{MOS}(n)_i = a_i \cdot e^{-b_i \cdot n} + c_i, \quad \forall i = 1, \ldots, 5$. Each of the 5 MOS functions $\{a_i, b_i, c_i\}$ corresponds to a different fraction gap λ_i (e.g., $\lambda_1 < 5\%$, $5\% < \lambda_2 < 10\%$, etc.). Figure 2(b) depicts the corresponding QoE estimation model. All the aforementioned techniques are based on YouTube as video streaming service, but can be easily extended to other HTTP-based streaming applications, simply by recomputing the player's buffering thresholds.

4. EVALUATION AND VALIDATION

Figure 3 reports some validation results of YOUQMON. Results correspond to 386 YouTube videos streamed from `youtube.com` through a bottleneck link of controlled capacity (from 128kbps to 20Mbps). Figure 4 depicts the characteristics of the corresponding videos, in terms of 4(a) video duration, 4(b) video size, and 4(c) VBR. Figures 3(a) and 3(c) show that the number and total duration of stallings per video computed by YOUQMON are highly consistent with the stallings measured at the YouTube player. Figure 3(b) shows that for 131 videos, the number of stallings is zero and the absolute difference between the estimated (n_e) and the real (n_a) number of stallings is 0. For the 255 remaining videos, the relative difference $\frac{|n_e - n_a|}{n_a}$ is still 0 for 30% of the cases, and below 15% for about 90% of the videos. Hence, for more than 93% of the 386 tested videos, the estimation is either exact or there are errors for $n_a > 6$. According to the QoE model depicted in figure 2(b), MOS differences for $n > 4$ are negligible, showing that YOUQMON is actually performing highly accurately in the validation dataset.

To validate the QoE estimation properties of YOUQMON, we replay one day of the network packet traces captured in the field trial study conducted in [2], for which the MOS values declared by the users are provided as ground truth. Figure 5(a) compares both the declared MOS and the YOUQMON MOS values for 16 different videos which experienced different stalling patterns in the field trial. Obtained results are very accurate and close to the MOS values actually declared by the participants, with only minor differences at the edges of the rating scale, which arise from rating saturation effects (e.g., in the field trial, ratings for 0 stallings correspond to MOS values around 4.5, while the model gives a MOS value of 5 in such cases) and are therefore not an issue.

Finally, figure 5(b) depicts an histogram on the number of reported tickets and the total played seconds of YouTube videos at the different estimated QoE levels, for one hour of real traffic monitored at the live 3G network. These results show that using YOUQMON it is actually possible to have a clear view of the performance of the mobile network as regards the satisfaction of the customers consuming YouTube videos. As reported by the charts in figures 5(c) and 5(d), the resulting YouTube QoE in this network is excellent (i.e., MOS = 5) for about 90% of the issued tickets and of the video time consumed during the analyzed hour. For 9% of the issued tickets and 4% of the total video time, the quality achieved was average (i.e. MOS = 3.4 in this case). Regarding bad quality events, YOUQMON can not say whether bad quality events come from problems on the network or in any other part of the end-to-end path (the customer terminal, the YouTube servers, a bad SNR, etc).

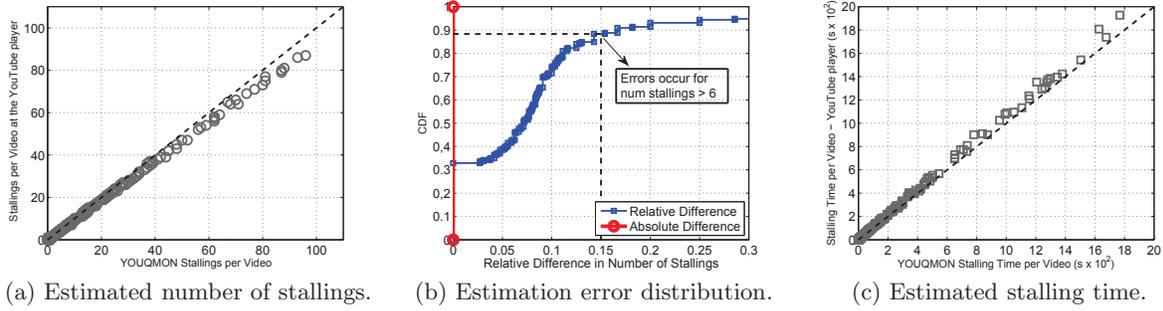| (a) Estimated number of stallings. | (b) Estimation error distribution. | (c) Estimated stalling time. |

Figure 3: YOUQMON validation. Estimated (a) number of stallings, (b) distribution of errors, and (c) duration of stallings for 386 YouTube videos streamed from `youtube.com` through a bottleneck link.

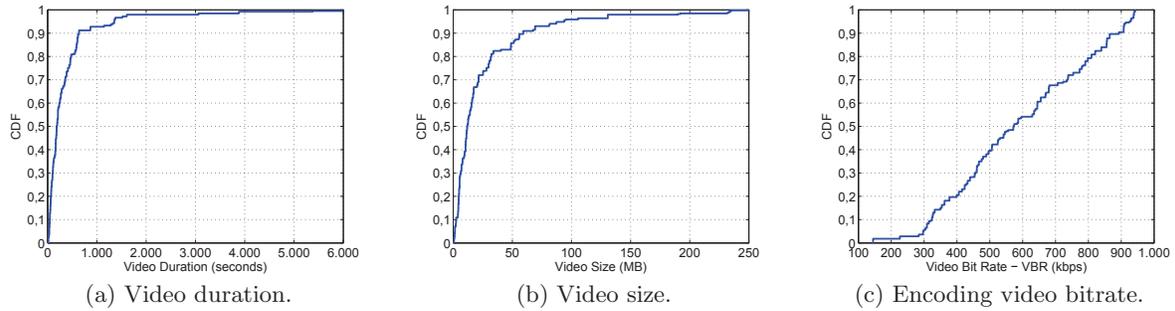| (a) Video duration. | (b) Video size. | (c) Encoding video bitrate. |

Figure 4: Characterization of the videos used in the validation of the stalling detector. More than 80% of the videos are shorter than 10 minutes, 90% are smaller than 60 MB, and the average VBR is 590 kbps.
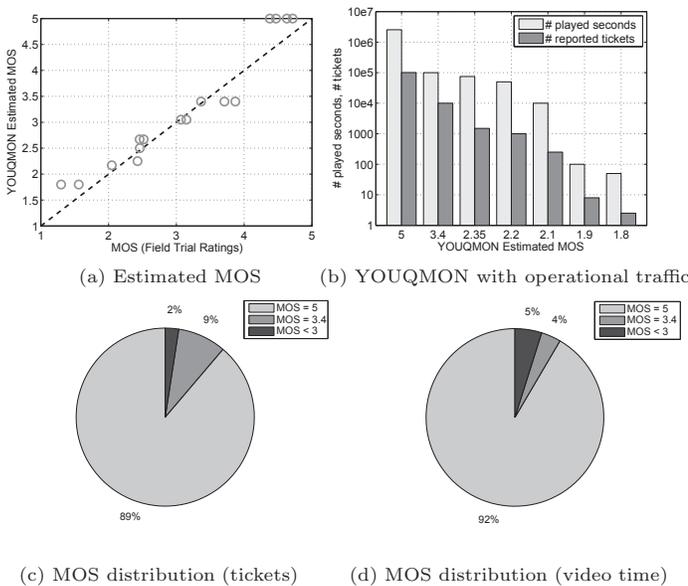

| (a) Estimated MOS | (b) YOUQMON with operational traffic |

| (c) MOS distribution (tickets) | (d) MOS distribution (video time) |

Figure 5: System validation in a mobile network.

5. CONCLUSIONS

In this paper we have studied the problem of YouTube QoE real-time monitoring and assessment in mobile networks. We presented YOUQMON, a system capable of giving concrete real-time indications on the performance of a mobile broadband network regarding the experience of the customers watching YouTube videos. In particular, we have studied the problem of how to extract YouTube performance indicators related to the QoE perceived by end-users, relying exclusively on packet-level measurements. To sum up, we have shown through YOUQMON the potential and feasibility of doing real-time QoE monitoring in services such as YouTube.

6. REFERENCES

[1] T. Hossfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, and P. Tran-Gia, "Quantification of YouTube QoE via Crowdsourcing ", in *MQoE*, 2011.

[2] P. Casas, A. Sackl, S. Egger, R. Schatz, "YouTube & Facebook Quality of Experience in Mobile Broadband Networks", in *IEEE QoEMC*, 2012.

[3] B. Staehle, M. Hirth, R. Pries, F. Wamser, D. Staehle, "YoMo: A YouTube Application Comfort Monitoring Tool", in *QoEMCS*, 2010.

[4] M. Eckert, T. Knoll, "An Advanced Network based Method for Video QoE Estimation based on Throughput Measurement", in *Proc. EuroView*, 2012.

[5] R. Schatz, T. Hossfeld, P. Casas, "Passive YouTube QoE Monitoring for ISPs", in *FingNET*, 2012.

[6] F. Ricciato, "Traffic Monitoring and Analysis for the Optimization of a 3G Network", in *IEEE Wireless Communications*, vol. 13(6), 2006.