This is the accepted version of a paper presented at ICIN 2016, March 2016, Paris.

Citation for the original published paper:


N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:bth-15636
Quality of Experience on Smartphones: Network, Application, and Energy Perspectives

Selim Ickin¹, Markus Fiedler²,
Blekinge Institute of Technology, Sweden²

Email: selimickin@gmail.com, markus.fiedler@bth.se

Abstract—For service and mobile operators, it is important to monitor and keep high user engagement levels. Quality of Experience (QoE) on video streaming applications is an important engagement measure for video consumer customers. In this paper, video QoE (with the focus on stalling events) is studied from network, application, and energy perspectives with various instrumentations on a smartphone. This enables the understanding of inter-relation between the perspectives and also how they influence the video QoE. Results show that packet delay variation and the maximal burst size in the network level; inter-picture time (picture delay) in the application layer; and also fluctuations in the energy consumptions are strong indicators for QoE. We show, via extensive QoE and energy measurements on smartphones that, based on the choice of streaming protocol, energy consumption can be reduced or increased in the case of stalling events during a video stream.

Index Terms—QoE, QoS, Energy, Mobile, Video

I. INTRODUCTION

Mobile video streaming traffic has exceed 50% of the world’s mobile data traffic in 2012, and it is expected to be three-fourths of the world’s mobile data traffic by 2019 [1]. Thanks to the 4G and beyond radio access technologies, which have given rise to the perceived quality of video streaming applications and services. The degree of delight or the annoyance of a user for a particular video streaming service is named as the Quality of Experience [2]. There is high competition amongst the operators with the aim to enable the highest QoE levels on the used video streaming services, which is important to increase revenues. QoE is related to many influential factors including the network, application, and energy on the end user device. The influential factors of QoE from different perspectives as well as the inter-relation amongst them needs to be well understood, in order to actuate, often on the network level by ISP’s, to obtain high QoE levels.

In this summary paper, we present summary of important findings in our previous work [6] [7] [4] [9]. Some of the influential factors in smartphone based video QoE are studied from different perspectives such as network, application, and energy. We also discuss the inter-relation in between different layers in the Internet stack. As an analogy, one can imagine a heavy traffic jam at a highway during a peak hour such that the vehicles consume fuel at a stand-still state. This is a situation where the cars don’t move or barely move, and at the same time consuming fuel as the engines are still running. Same with a video streaming application on a smartphone in communication networks domain; the video streams are stalled occasionally due to the Internet packet latency in the mobile network caused often by coverage issues or heavy load, e.g., high number of users, on the mobile network cell. This might eventually increase energy consumption. This paper focuses on the stalling, as previous research shows that it is one of the most important influential factor on video QoE [13]. In this paper, we study the energy perspectives of QoE together with the network and the application as all of these have complex inter-relationships between each other.

The paper is structured as follows. In Section II, we present some of the important influential factors belonging to aforementioned perspectives. Sections III-V present the instrumentation of QoE, for 3G-based video streaming, from the network, application, and the energy perspectives, respectively. The conclusive remarks from various studies relating the QoE to network, application, and energy are given in Section VI. Section VII concludes the paper.

II. INFLUENTIAL FACTORS ON CELLULAR BASED VIDEO QoE

The degradation of QoE levels in cellular based video streaming is often, although not necessarily, caused by degradation of QoS level in the radio network level. A high packet latency caused by a heavy load in the network cell influences the delivery of video data, e.g., the received throughput on the smartphone. For example, the choice of a transmission protocol might influence the energy consumption when there is a problem in the network. If a video packet has not been received within a particular time window, then the video packets are re-transmitted from the source video server, which in turn might cause video packets to be accumulated in large queues at the radio link, e.g., in the basestations. This impacts QoE indirectly in many aspects including user’s monthly data usage offered by the network operator, the presentation of the video content to the user in the video streaming application, and the energy consumption of the device due to increased duration of the cellular network module’s active state. Increased mobile data usage caused by re-transmission of video data impacts QoE as it increases the monthly data
cost of customers. The presentation of the video content to the end user through the user interface, i.e., the device screen, is also interrupted and manifested as stalling events which in turn degrades the video QoE of users. The increased energy consumption indirectly affects the QoE, especially in battery powered mobile devices, as the increased energy consumption reduces the operation time of a device with the increased drain of battery voltage. Thus, saving energy on smartphones can both increase the operation time of a smartphone, and also contributes in greening the network. A word cloud in Fig.1 is constructed based on the frequency of the words obtained from 29 users via 376 expressions in Day Reconstruction Method (DRM) weekly interviews (conducted in the lab) as well as 430 entries in the online survey [4]. DRM is a method to help users to provide a feedback on the perceived experience within the last 24 hours at each recalled activity on the smartphone. The most frequent keyword is battery consumption on smartphones which is followed by many other including mobility, Internet performance (e.g., ‘slow’, ‘freeze’), camera, Flash Player. The coding and grouping of the words are performed by two researchers with expert knowledge, with a 90% aggregation rate. The inter-relation between the influential factors on QoE is depicted in Fig. 2. The indicative metrics at the network, application, and energy aspects are discussed next.

III. QoE STUDY IN THE NETWORK

We study the video QoE with respect to the metrics measured directly at the network level. The change of QoE caused by a change of Quality of Service (QoS) depends on the actual current level of QoE [5]. The IQX hypothesis, in relation to ON/OFF video models, shows QoE as an exponential function of QoS. We have implemented and deployed a measurement Linux kernel module at the Android device. This measurement module records the timestamp when the video packets arrive at the smartphone terminal. In addition, the measurement module is also deployed on the media server such that the timestamps when the video packets leave the media streaming server are also recorded. This is implemented on top of a User Datagram Protocol (UDP) tunnel such that at the server side, metrics such as the sequence number and the time stamp of the departing packets are appended to the UDP tunnel header and then sent to the receiving end. Once the packet is received at the smartphone terminal, these two metrics are decapsulated. This way, the end-to-end delay as well as delay variation during a video stream is measured, and then matched to the video QoE. QoE is measured via 5-level Mean Opinion Score (MOS) scale. The communication between the kernel and the user space are done via a UDP socket communication. One-way-delay of one packet \( D_n \) is calculated by the subtraction of the departure timestamp \( T_{S,n} \) from the arrival timestamp \( T_{R,n} \) obtained at both ends of the communication channel as shown in Eq. 1.

\[
D_n = T_{R,n} - T_{S,n} \text{ [ms]} \tag{1}
\]

Packet Delay Variation (PDV) values, and the user rating (UR) values are given in timeseries. It shows a user is reluctant to give high user ratings when the video quality improves, but immediately reacts with a poor user rating when the quality degrades. The benefit of using the Exponential Weighted Moving Average (EWMA) techniques on human perception statistics is further studied in the scope of memory effect as the current QoE of a user highly depends on the previous QoE [11] [12]. Inclusion of the remaining effects of the previously obtained outputs into the calculation of the current output is made possible by the EWMA approach (as computed in Eq. 3). \( PDV_{EWMA}(i) \) is the current (at \( i^{th} \) interval) exponential weighted moving average PDV, \( PDV_{EWMA}(i - 1) \) is the previous (at \( i - 1^{th} \) interval) exponential weighted moving average PDV, and \( PDV(i) \) is the current PDV. \( \alpha \) is typically set to 0.25.

\[
PDV_{EWMA}(i) = (1 - \alpha) \cdot PDV_{EWMA}(i - 1) + \alpha \cdot PDV(i) \tag{3}
\]

When obtaining the User Rating (UR), EWMA is used for computing the correlation of instantaneous user perception against the current and the previous QoS metrics. Thus, we...
imitated the human perception by using EWMA on PDV values [11].

We used a user interface similar to Fig. 4(a), and asked each user to rate the video quality by using one of the five buttons located in the user interface at own will while the video is being streamed. There were in total 15 participants in the study, and they were asked to watch the video as they have been watching it in daily life settings.

Various video packet queues on the way from the streaming server to the video player might cause the PDV. Moreover, the 3G operator contributes to increases in PDV by attempting to compensate packet loss by retransmissions. In addition, TCP, while trying to recover from lost packets, might amplify the PDV. This might yield bursty packet traffic causing stallings. For this reason, we study the packet delay variation in an other metric called Maximal Burst Size (MBS). This measures the amount of packets being received at a smartphone terminal in a short time interval. We assumed that this way the abnormal bursty behaviour of the video streaming traffic via 3G can better be captured.

First, we observe that the UR obtained at a given time is strongly impacted by the previously measured PDV values as expected. The power-law model fitted better (with approx. a 0.1 unit larger $R^2$ value) than an exponential one, which we think is due to the ON/OFF behaviour of a 3G-based video streaming yielding the bursty traffic. The goodness of the fit, the $R^2$ value for the power model, is improved by over 100% when the EWMA technique is applied. We obtained the models for the relation between the PDV and the MOS that is strengthen via EWMA as given with Eq.4.

$$UR = -9.10 \times PDV_{(ms)}^{0.08} + 16.18, \quad (4)$$

The relation between the UR and the MBS is also studied, and observed that the peaks in the number of transmitted packets per given time interval have a negative impact on the UR. This can be explained by the fact that large MBS values indicate a ON/OFF behavior. In this case, when there is no available bandwidth the packets are queued, and then flushed all at once causing a stall in the video streaming application, as the application is not able to process large volume of data within a particular time interval. Eq.5 shows the obtained model with respect to the MBS and the MOS with a better ($R^2$=0.78) goodness of fit as compared to Eq.4 ($R^2$=0.68).

$$UR = 59.96 \times MBS_{(packets/ms)^{-0.036}} - 51.71 \quad (5)$$

In order to find out the influential factors on video QoE, we investigate at a level in the Internet stack that is closest to the user. Typically, the user interface level is where a particular service is sensed and experienced by a human user.

IV. QoE STUDY IN THE APPLICATION

In the application, we consider the user interface, the point where the user interacts with a video streaming application/service. Measurement points are deployed at relevant parts of the open-source VLC video player application code base such as the recording of timestamps when the video frame is rendered and displayed on the video screen. This modified version is called VLQoE. Then, the inter-frame (or inter-picture) time during a video stream are are computed. The deviation of the inter-picture times are then matched to the perceived video QoE. We measured and quantified QoE with opinion scores that are collected at user’s own will while the video is being streamed. We conducted the measurements on an Android device.

The users in this study are asked to watch a 250 seconds long video clip that consists of various scenes including racing scenes of sailing boats. The video clip contains a sequence of 6251 pictures and is encoded with a nominal frame rate of 25 fps with a bitrate of 1000 kbit/s. The same video is watched twice (first with RTSP then with Hypertext Transfer Protocol (HTTP)) by each user on a smartphone (with a video screen size of $196 \times 117$ pixels, which is provided to them. Each user was encouraged to rate the temporal quality based on the five-level MOS scale, while pressing one of the five user rating buttons at her/his own will during the playout. In addition, a ‘freeze’ button is horizontally placed on top of the five buttons at the user interface. The user interface used in the experiments is given in Fig. 4(a).

30 subjects performed the study at various location in Karlskrona, Sweden by using exactly the same smartphone. In total, 60 (30 users x 2 protocols) user experiments are conducted. We asked the users to hold the smartphone at a comfortable distance with convenient illumination level, i.e., at a familiar physical context as in daily life. This way, we imitated peoples’ natural daily life settings in the subjective tests. The video was muted such that the users could focus only on the visual freezes. A visual freeze can be measured at the user interface with an inter-picture time $D_p$, as the time gap between two consecutive pictures displayed on the smartphone screen. This is calculated in Eq.6. $T_p(k)$ is the timestamp when the $k^{th}$ picture is displayed on the smartphone screen. An illustration of the $D_p$ metric during a video stream is given in Fig.4(b).

$$D_p(k) = T_p(k) - T_p(k - 1) \quad (6)$$
The $D_p$ values in between two consecutive user indications, $T_i$ are considered. The distribution of the inter-picture time values for the user ratings $1 - 5$, and freeze indications collected from all users is given in Fig. 5. Mean $D_p$ values are 152 ms, 282 ms, 321 ms, 768 ms, and 831 ms, for ‘UR 5’, ‘UR 4’, ‘UR 3’, ‘UR 2’, ‘UR 1’, and freeze, respectively. Van Kester et al. [10] states the acceptable freezing duration as 360 ms which also confirms our results considering an acceptable UR as 3.

By using the VLQoE tool, we assumed that a 3G video stream follows a two-state ON/OFF model. The ON (smooth playout) and OFF (freeze) states are set based on the $D_p$ metric. We considered the state of the art 100 ms as the maximum tolerance threshold for a user to feel that a system reacts instantaneously [3], and then $D_p$ values less than 100 ms are set to an ON state; while the $D_p$ values higher than 100 ms are set to an OFF state. The mean $R^2$ values were calculated as 0.81 and 0.93 for ON and OFF states, respectively. Next, the Maximum Likelyhood Estimate (MLE)s for the durations of ON and OFF states for all 58 iterations are calculated, then the mean of all MLE values are obtained. This model is applied for controlled local-based video streams in the user tests as described in Section IV. The mean MLE of ON and OFF durations is calculated as 9.7 s and 642 ms, respectively.

V. ENERGY AS AN OBSERVABLE METRIC FOR QoE INDICATION

Energy, particularly the remaining battery level of a smartphone, amongst many other influential factors, is one of the most important ones that influence the overall QoE of a smartphone user [4]. And energy is highly consumed in video streaming applications as they are both high bandwidth demanding, keeping the network module at active state for long durations, and in parallel necessitate high CPU utilization, for processing/rendering the multimedia packets, and also good-enough screen light, for a clear presentation of video content on a battery-powered smartphone display. For this reason, it is important to understand the energy consumption patterns during a video stream and how these patterns are influenced when there are video quality issues such as stalling events.

By using the VLQoE tool, the inter-picture time during a video stream is measured. In parallel, Monsoon energy measurement tool is used to measure the total instantaneous power consumption of a smartphone terminal. Then, the relation between the power measurements and the freezes are observed. Fig. 6 illustrates a snapshot, during a video stream, the relation between the inter-picture time and the total power consumption of smartphone. Phase 1 is the low power state, can also be called as the initial rebuffering state where no pictures are yet displayed on the screen, of the video stream. Phase 2 is the steady state playout state. It is observed that the high inter-picture time causes a slight reduction in the power consumption values. The reason for this is that the video stalls, as there are no packets to present on the smartphone screen. Thus, there is a direct relationship between the inter-picture time and the duration of the power consumption staying at a rather lower level. Fig. 6 is a high level sketch of a smartphone’s power consumption pattern when there is a stalling event during a video stream. We hypothesize that...
actually freezes can cause either an energy saving or an energy waste depending on the transmission protocol. In the case of a freeze during a video stream, if there is a packet retransmission such as if the video stream is established over a Transport Control Protocol (TCP) connection, then all the video content is shown to the user with an extended view duration. During the freeze time, the energy is being consumed due to a set of factors including CPU, AMOLED display, low power state of 3G data module. Thus, the energy consumed during the stall duration is considered as wasted. Therefore, a smoother video playout might cause a lower total energy consumption and a higher QoE. In contrast, if there is no retransmission to compensate the loss packets, such as in video streams established over UDP, then there is a picture jump, which causes some content to be skipped causing the total video view duration unchanged. The energy drop during a freeze duration can be considered as the saved energy.

On the other hand, the freezes influence the QoE of a user during a video stream. Thus, in the case of a video transmission with skipped pictures, there is a tradeoff between the energy saving and the MOS. It is important to find out the maximum saved energy without impacting QoE. For the case of video streams, where there is no picture jump, the waste of energy increases with the duration of a freeze. Thus, for the latter scenario, it can be said that the better the streaming quality, the less energy is wasted. The two scenarios are sketched in Fig. 7. Next, the amount energy waste and the energy savings for the two scenarios are computed by considering QoE models. The procedure that is followed during the video experiments is given in Fig. 8. Each user is asked to watch three versions of the same video content in a random order, and are asked to give a MOS score at the end of each video. The film clip was three minutes long, with a 6Mbit/s bitrate and 25 fps, and converted into MP4 multimedia format.

The video is streamed via the local drive of the smartphone to enable controlled experiments. We emulate the ON/OFF exponential model obtained in Section IV, such that the freeze durations are exponentially distributed over the video session. The mean OFF duration is set to 2 s with varying mean ON durations 4 s, 8 s, and 16 s. The first version of the video is the original one that has no temporal distortions (scenario 1); the second version contains freezes and the video pictures are not skipped (scenario 2); the third version also contains freezes and the video pictures are skipped (scenario 3). The three versions of the videos are shown to the user in a random order, and in between each video session, the user is asked to rate the video quality of the previous film clip. In Scenario 2 and 3, the mean OFF duration is set to 2 s. After the user registers the MOS score, the user is asked to watch a gray screen for 15 seconds on the smartphone to reduce the memory effect. In total, 60 users are involved in the study. Next we computed the OFF probability, \( P_{\text{OFF}} \) as the ratio of the mean OFF state duration to the sum of the mean ON and OFF state durations. Then, the relation in-between the \( P_{\text{OFF}} \) and the MOS scores are computed.

In total, it was observed that, with the presented settings in the experiment, there was no statistically significantly difference in QoE regardless of the video freeze is associated with or without skipped video content. The important factor that influences the QoE is whether or not a freeze happens. Thus, we merge the data for scenario 2 and 3, i.e., the scenarios involving freezes regardless of the fact that the pictures are skipped or not. When the \( P_{\text{OFF}} \) values are fitted to the MOS scores, an exponential relation is obtained in Eq. 7 with a \( R^2 = 0.73 \).

\[
MOS = 4.59e^{-3.44 \cdot P_{\text{OFF}}}
\]  

Eq. 7 can further be used to calculate the relation between the energy saving or energy waste as the energy saving or the waste is highly related with the ratio of the total freeze duration to the whole video duration. If this ratio is assumed to be equal to the \( P_{\text{OFF}} \), then Eq. 8 is obtained.

\[
MOS = 4.59e^{-3.44 \cdot \frac{T_{\text{freeze}}}{T_{\text{video}}}}
\]  

The power saving during a freeze, \( P_{\text{saving}} \) is measured to be 185 mW, and the total freeze duration during the video stream, \( T_{\text{freeze}} \), can be further substituted with the \( E_{\text{saving}} \) yielding a final model as in Eq. 9, where the total video duration, \( T_{\text{video}} \), is three minutes in the experiment setup. With this tradeoff model...
between the MOS and the energy saving, at most 4.25J can be saved while keeping the MOS level at 3.

\[
MOS = 4.59 \cdot e^{-18.59 \frac{P_{\text{onset}}}{\text{video}}} \tag{9}
\]

The power consumption during a freeze, \(P_{\text{freeze}}\) is measured to be 728 mW. Applying similar calculations to model the relation between the MOS and the energy waste, Eq. 10 can be obtained showing that the MOS degrades further with the increase in the energy waste caused by a freeze.

\[
MOS = 4.59 \cdot e^{-4.72 \frac{E_{\text{waste}}}{\text{video}}} \tag{10}
\]

The amount of energy saving in a three minutes long video stream is insignificant for a commercial single smartphone with battery capacity of 9.88 Wh. However, as more users spend more and more time on video streaming on mobile devices every day, the total amount of energy saving is expected to increase further more.

VI. CONCLUSIVE REMARKS

A set of QoE measurements are instrumented based on the network, application, and energy perspectives. We have presented the relation between the energy and the QoE and showed examples of video streaming scenarios for tradeoff and win-win. Packet delay variation and the maximal burst size metrics measured at the network packet level are studied as strong indicators of QoE. Although an exponential model represented the relation between QoE and QoS well, a power-law model is found to yield slightly better results. One explanation of a power-law model can be the bursty video delivery at the radio link. This includes long time intervals without any throughput received at the receiving video streaming client followed by a bursty delivery of packets. In this case, the long-delayed packets are delivered all at once (with a very small packet delay variation) causing stalling events during a video streaming application.

The stalling events during a video stream is studied at the application layer via quantifying a stalling event with an inter-picture time. This is instrumented by modifying the open-source VLC player for Android OS. Inter-picture time of a 3G-based video stream can be represented via a two-state exponential ON/OFF model with a mean OFF value of around 600 ms and mean ON value of around 10 s. Based on the results obtained from the user study, the mean \(D_p\) (inter-picture time) values are 152 ms, 282 ms, 321 ms, 768 ms, 831 ms, and 1289 ms for ‘UR 5’, ‘UR 4’, ‘UR 3’, ‘UR 2’, ‘UR 1’, and ‘Freeze’, respectively.

Throughout extensive energy measurements on the smartphone, we identified potential scenarios where energy saving might be possible and we recommended approaches to increase energy saving while maintaining QoE. It has shown, for TCP-based streams, that a lower total energy consumption can be achieved with a smoother video playout causing in parallel a higher QoE. In contrast, for UDP-based streams, there is a tradeoff between the QoE and the energy saving, i.e., the stalling events both decrease QoE and the energy consumption.

VII. CONCLUSION

In this paper, we have presented various studies on video Quality of Experience (QoE) from different perspectives such as network, application, and energy. We present QoE models as a function of Packet Delay Variation (PDV) at the network level; and \(P_{\text{OFF}}\) (freeze or OFF probability) measured at the user interface. Next, we leverage the QoE models to enable energy saving on the smartphone. We conclude that a smoother video yields a better QoE, and less energy consumption for retransmission-based video streams. We present QoE models that show the relation between the energy with respect to the stallings at the user interface.

Saving energy is possible with a better video streaming quality. The QoE of a video streaming application/service highly depends on the network level QoS metrics including the packet delay, throughput, and more. Thus, the most obvious way of monitoring and preventing poor video quality is via QoS management in Internet Service Provider (ISP)’s, e.g., with smart scheduling and increasing bandwidth. However, due to the wide variety of video streaming applications with different streaming characteristics and over the top streaming protocols, it is hard to pinpoint a low QoE level by relying on the QoS which is mainly the metrics collected in the network level. For example, a high packet delay, or a poor throughput in the network link does not necessarily result in a video freeze if the video buffer length is large enough to tolerate the packet latencies. The problem reveals itself better for video applications that have low tolerance to the end-to-end packet delays such as online gaming and remote surgery. As future work, a collaborative data analytics and machine learning techniques can be suggested to study many cross-layer metrics simultaneously to predict quality degradations in advance and to actuate timely robust decisions to improve QoE and to save energy for smartphones at the same time.

REFERENCES


