Packet Loss based Quality of Experience of multimedia video flows

René Serral-Gracià*, Yue Lu†, Marcelo Yannuzzi *, Xavier Masip-Bruin *, Fernando A. Kuipers †
* Centre de Recerca d'Arquitectures Avançades de Xarxes (CRAAX) (UPC), Spain
Email:{rserral, yannuzzi, xmasip}@ac.upc.edu
† Delft University of Technology (TU-Delft), The Netherlands
Email:{Y.Lu,F.A.Kuipers}@tudelft.nl

***Abstract*—Quality of Experience assessment of multimedia services is a hot topic in current research. This is motivated by the broad interest that both users and content providers show in having a successful multimedia experience.**

In this line, there has been some proposals to assess the QoE over multimedia video flows. However, most of these proposals are complex to implement, and require to compare the received stream with the original, which is unpractical in most scenarios.

In this paper we present a novel method to compute the user's perceved video quality. Opposed to other works, our method solely uses the video received at destination in order to assess its quality. To accomplish this, we base our quality estimation in the analysis of the experienced frame losses and the application buffers.

To demonstrate the accuarcy of our solution, we compute the experienced quality of a video streaming application under different network disruptions, comparing the results obtained by our solution with the well known video quality metric Peek-Signal to Noise Ratio.

I. Introduction

Nowadays the Internet is evolving to a content oriented network, and as a such, one of its uses is to deliver various sorts of multimedia content to its users. This content can be provided via different services, e.g., through HTTP streaming, through subscriber based IPTV services, through Peer-to-Peer (P2P) TeleVision (P2PTV) broadcasting, etc. These services can have different billing policies, such as predefined monthly fees, pay-per-view, while others can be provided with no direct cost to the user. However, despite its source, all users strive for a quality multimedia experience. Consequently, many efforts are invested by content providers to assess the quality of a particular multimedia service.

The issues with the assessment is that multimedia services are bound to different constraints than regular traffic, specially in terms of network performance (the so called Quality of Service – QoS) and subjective quality perceived by the users (also known as Quality of Experience – QoE). To cope with this, current research efforts are centered in proposing mechanisms for QoE assessment.

QoE measures a broad range of parameters, from the user satisfaction in watching a movie, to the user satisfaction in the responsiveness of channel hopping in a Television distribution system. However, all the different aspects regarding QoE evaluation have a common point, i.e., the user satisfaction measurement. Given that QoE is based on user satisfaction, it is clearly bound to a certain *subjectiveness*. Despite of this, QoE measurement is very important to content providers, in particular, when deploying a new service, when planning a network upgrade, or when deciding the billing policies for a service. Accordingly, the content providers are determined to *objectively* assess the user perspective of a given service.

Current solutions to the QoE assessment only consider one single aspect of the quality (e.g. video quality, or response time, ...), while human perception of quality is composed by several conditions, which depend on the specific service. To fill this gap, in this paper we introduce a framework for QoE assessment of multimedia content, named Profile-Based QoE Assessment Framework (PBQAF). The goal of this framework is to provide a stable platform for continuous monitoring the QoE of multimedia services.

Given its design, our framework can be applied to any context where multimedia services are involved, permitting to assess the QoE in different environments (e.g. P2PTV, VoD, video-conferencing, ...), as it uses cross layer information that can be used to effectively assess the perceived user experience. In fact, we use the PBQAF to assess the QoE of state-of-the-art multimedia streaming solutions such as P2PTV, where the classical client/server paradigm is changed to an all-against-all environment, and assessing the multimedia streaming quality poses a whole new set of challenges, which are not addressed yet in the literature. In particular, we objectively study a case of user satisfactibility in the deployment of a P2PTV application (i.e., SopCast) at 137 different nodes within the PlanetLab network. It is worth mentioning that, as far as we know, this qualitative study has never been done before.

The rest of the paper is structured as follows. In the next section we present the different related works in QoE, focusing on P2PTV environments. We continue the discussion in Section III by presenting our general framework and its different building blocks. In Section **??** we detail the specific methodology we use for the QoE assessment. Later, in Section IV, we evaluate our framework within the PlanetLab network, and finally in Section V we conclude and outline our future research.

This work was partially funded by MCIN (Spanish Ministry of Innovation) under contract TEC2009-07041 and CIRIT (Catalan Research Council) under contract 2009 SGR 1508.

II. Related Work

QoE metrics definition and evaluation is a challenging topic where many research efforts are centered. In general, when assessing the QoE of a new video codec, or a new network protocol, the common procedure is to perform a survey of user satisfactibility. This involves gathering some individuals, who, in the case of video quality, have to score the perceived quality of a set of videos under different controlled conditions (usually with network or video disruptions). Such score is known as the Opinion Score, which is a number ranging from 5, i.e., perfect perceived quality, to 1, meaning total lack of quality. Since the experiments involve a set of individuals, the survey's outcome is the subjective Mean Opinion Score (MOS) [1].

MOS was originally devised for voice services. Given its usefulness, in [2] ITU-T proposes a mechanism to objectivize MOS computation for voice services with the E-Model. Lately, in works such as [3], by using the Peak Signal-to-Noise Ratio (PSNR), the authors propose a MOS compatible scale for video services. However, the main drawback of using PSNR in a real scenario, is that the quality assessment requires the original video. This clearly limits the applicability of PSNR in a distributed scenario such as the Internet. In particular, this mechanism is not applicable in a continuous QoE assessment framework, as the one proposed in our paper. Moreover, the PSNR comparison cannot determine non video-quality related factors, such as video start-up time, level of interactivity, etc.

FIXME: Canvia lo de baix per a ser només de video i en un sol punt d'anàlisi Even if such efforts are invested in deploying P2PTV infrastructures, there is not, to the best of our knowledge, any existing solution that can assess the QoE in such environments. In particular, existing solutions to QoE measurement and network planning are often designed to operate with a specific type of service in traditional transmission systems, e.g., [4] studies QoE under VoIP services, while [5] focuses on IPTV solutions, and [6] reviews QoE assessment for MPEG-4 video-streaming. On the contrary, our approach aims at providing a more general solution for accurate QoE assessment, which is independent of the underlying network technology, and of the specific media content.

III. Video flows Quality of Experience computation

This section describes the main contribution of this work, here we present the methodology used for the QoE computation. FIXME: lo de baix

1) Direct metrics: The direct metrics are computed differently depending on which acquisition mechanism is used: *Payload* or *Run-time analysis*.

In the case of Run-time analysis acquisition mechanism, to assess the video quality, we identify three different frame status, *i)* Correct, *ii)* Disrupted, *iii)* Lost. In this case, Correct stands for no packets have been lost and all arrived on-time. Disrupted means that at least one packet of the frame arrived on time. And finally a frame is considered Lost when no packet belonging to the frame arrived on time to the end-user.

Besides, from the application point of view we can infer, depending on the frame types (I, P or B), the length of the disruption. To compute this duration we assume the insights provided in [7].

Therefore, the frame quality can be easily mapped to numerical values by using the conversion rule presented in Table I, where δ $(0 \leq \delta \leq 1)$ is a parameter specified by the profile, that defines the importance of the video quality for the given service. The lower δ the worse will treat the system the sporadic packet losses.

Frame	Quality
Correct	1
Disrupted	δ
Lost	0

TABLE I
Frame quality conversion table

In the case that it is not possible to access run-time information, the available information is limited to packets, without any knowledge about frame types. Hence, the assessment for a particular frame f in this case is performed by Equation 1 as follows:

$$q(f) = M\left(PLR(f)\right) \tag{1}$$

where M is the mapping function defined by the profile, and $PLR(f)$ is the packet loss ratio of the frame f. A simple mapping function for M can be: $M(x) = 1 - x$. Hence when there is a high packet loss ratio we have low quality.

Therefore, the final quality of the video stream is the set of particular frame qualities defined by $Q = \{q(f_1), \ldots, q(f_n)\}$, where $f_{1 \ldots n}$ is the set of all the frames in the video stream. Analogously to MOS or PSNR we can compute the average to have a single quality value for the video $\overline{Q}$, or per time interval.

IV. Experimental Evaluation

Up to this section, the proposed PBQAF has been deeply described. In this section, we evaluate it. Hence, in order to evaluate PBQAF we deployed in the PlanetLab network a set of 137 nodes with SopCast, a P2PTV application. These nodes are composed by one server, located at the TU-Delft premises, and 136 clients located in several countries within the PlanetLab network.

The testing was performed by streaming with SopCast a 20 minutes long video from the server to the clients. Such video was a documentary with a 480 x 384 resolution and bit rate of 410kbps.

The accuracy analysis of PBQAF in QoE Assessement is issued by the following steps, first we setup a video capture platform, for dumping to disk the received video in all the reception nodes. Second, at the same edge nodes, we store the network traces by running `WireShark`. And finally we start the video streaming in the server and the clients collecting all the data for later processing.

To ease the comprehension, we analyze the direct and indirect metrics separately. In particular, the different aspects we study are, the user perceived video quality, the Start-up time, the blocking periods due to jitter, and most specially the different degrees of accuracy we can obtain by using *Run-time* and *Payload* analysis acquisition mechanisms.

We remove from the results all the nodes that experienced some kind of misbehavior, such as clock synchronization, CPU overhead, or traffic collection problems, leaving a total of 95 nodes for the analysis.

A. Results

With the direct metrics analysis we aim at showing the video streaming user's perceived quality. And compare it with the different acquisition mechanisms presented in the paper. We assess the accuracy of our framework by comparing the *"Perfect Knowledge"* obtained from our captures with the results measured by our system. This comparison is performed by correlating the our measurements with the reference PSNR. Moreover, we highlight the differences in accuracy depending on the two main acquisition mechanisms, namely *Application* and *End-Point*.

In Figure 1 we show the obtained overall video quality and the correlation coefficient in ascending order for all the nodes of the test, for both the *Run-time* and *Payload* analysis acquisition mechanisms. Such correlation coefficient is computed in relation with the PSNR that we take as reference. The analysis was performed with $\delta = 0.5$ for *Run-time* and $M(x) = 1 - x$ where x is the packet loss ratio for the *Payload* analysis.

As it can be noted, by having specific codec, and frame-type information, permits that the accuracy in the case of having *Run-time* information is always better. In fact, *Payload* analysis, tends to overestimate the quality because it this case it does not consider knowledge about the type of the dropped frames. Moreover, it cannot access the application's buffer status to improve the accuracy.

Looking closer at the results, we can see that our system has a correlation higher than 95% for more than 96% of the nodes, this is an accurate result with a very resource efficient technique for disruption detection, where we consider application run-time information such as buffer sizes and other such as frame types to estimate the effects of the loss periods. In the case of *Payload* analysis, the results are less accurate, we have more than 90% correlation for 70% of the tests. This reduction in accuracy is caused by the lack of information in this layer, where we considered packet losses and delay variations, but without accessing any application layer information such as frame-types, which could greatly improve the accuracy in this case. However, if the used encoding of the frames permitted it, we could improve the accuracy by accessing the frame types present in the packet's payload, but SopCast is a closed protocol which makes this task very difficult.

V. Conclusions

In this paper we presented a full fledged Profile Based QoE Assessment Framework. The main advantage of this system compared to other solutions is the fact that it can be easily adjusted to cope with many different environments.

As a proof of this, we deployed SopCast in a fairly large set of nodes within the PlanetLab network, and used our framework to assess the QoE of such a deployment. Our findings show the good accuracy in QoE assessment of our system, along with the detection of a few issues present in most P2PTV applications, such as unsynchronization and blocking.

Finally, as future lines of research, our goal is to further investigate the *Payload* mode of operation, specially in P2PTV scenarios, to improve the accuracy by gathering more application information to remove the requirement of having *Run-time* to obtain high accuracy levels.

References

[1] ITU-T Recommendation P.800. Methods for subjective determination of transmission quality, 08/96.

[2] ITU-T Recommendation G.107. The E-model, a computational model for use in transmission planning, 03/2005.

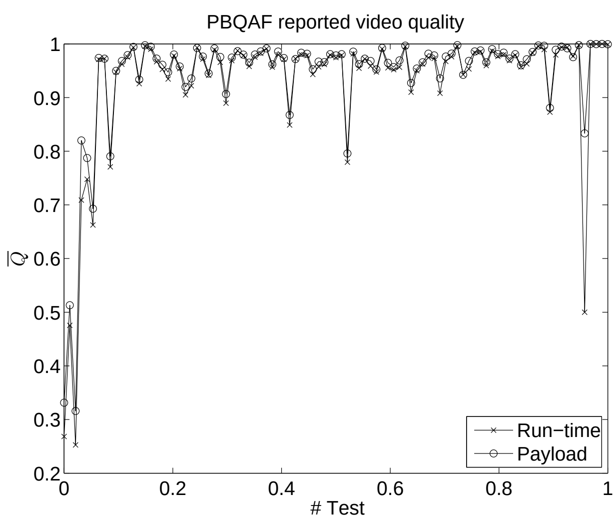
[3] Jirka Klaue, Berthold Rathke, and Adam Wolisz. EvalVid - A Framework for Video Transmission and Quality Evaluation. In *In Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 255–272, 2003.

[4] René Serral-Gracià, Loránd Jakab, and Jordi Domingo-Pascual. "Measurement Based Call Quality Reporting". In *2nd Workshop on Network Measurements in 32nd IEEE Conference on Local Computer Networks*, pages 997–1004, 2007.

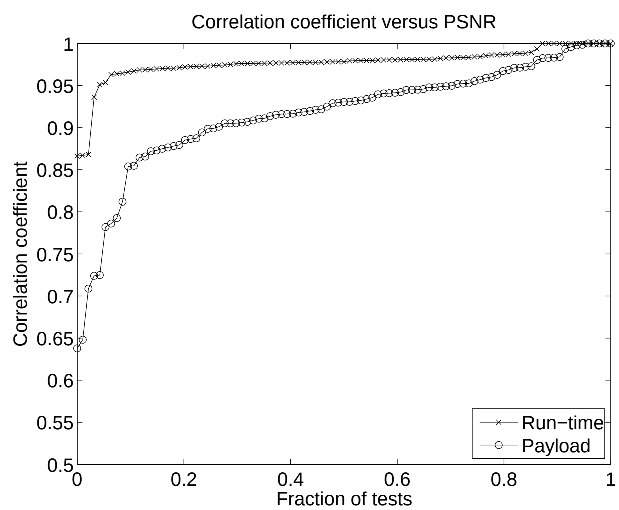
[5] Marcio Nieblas Zapater and Graca Bressan. A Proposed Approach for Quality of Experience Assurance of IPTV. In *ICDS '07: Proceedings of the First International Conference on the Digital Society*, page 25, Washington, DC, USA, 2007. IEEE Computer Society.

[6] James Gross, Jirka Klaue, Holger Karl, and Adam Wolisz. Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming. *Computer Communications*, 27(11):1044–1055, 2004.

[7] Jason Greengrass, John Evans, and Ali C. Begen. Not All Packets Are Equal, Part 2: The Impact of Network Packet Loss on Video Quality. *IEEE Internet Computing*, 13(2):74–82, 2009.



(a) Quality



(b) Correlation

Fig. 1. Video Quality comparison with Run-time and Payload analysis acquisition mechanisms