Live video-streaming evaluation using the ITU-T P.1203 QoE model in LTE networks

H.-F. Bermudez[1,2], J.-M. Martinez-Caro[3], R. Sanchez-Iborra[4], J.L. Arciniegas[1], M.-D. Cano[3]
[1] Dept. of Telematics, University of Cauca, Colombia
[2] University of Quindío, Colombia
[3] Dept. of Information Technologies and Communications, Universidad Politécnica de Cartagena, Spain
[4] Dept. of Information and Communication Engineering, University of Murcia, Spain

hfbermudez@uniquindio.edu.co, josem.martinezcaro@upct.es, ramonsanchez@um.es, jlarci@unicauca.edu.co, mdolores.cano@upct.es

***Abstract*— Nowadays, video-streaming services have become the most consumed service by end-users in mobile networks. This has led telecommunication operators to develop a series of networking mechanisms that allow to provide end-users with the demanded levels of quality. In order to evaluate the degree of user satisfaction, there are numerous methodologies that range from Quality of Service measurements (QoS), which are based on network parameters, to estimations of the Quality of user Experience (QoE), which is centered on users' perception. In this context, the ITU-T has recently released the ITU-T Rec. P.1203, which presents an objective model for estimating QoE in video-streaming services. In this paper, we present a functional implementation of the ITU-T Rec. P.1203 and its validation, as well as the guidelines for its application. As a practical case study, we evaluate the performance of a live video streaming service over an LTE network in terms of QoE using the ITU-T Rec. P.1203 model. The QoE outcomes from two different test-benches, real and emulated, and two TCP-based video transport protocols, DASH and RTMP, are compared. In addition, several propagation models are also tested in the emulated environment to find out the one that best matches real conditions based on QoE. From the results, we obtain that Nakagami-*m* with *m* = 5 is the most suitable propagation model to be used in the emulation of outdoor LTE scenarios and that the number of stalling events is the factor with the greatest importance in the QoE estimation provided by the model.**

***Keywords*— DASH, ITU-T P.1203, Live video streaming, LTE, QoE, QoS, RTMP.**

I. INTRODUCTION

Cellular-based communication technologies such as 4G are enabling end-users to access a plethora of cloud-based multimedia services such as YouTube, Facebook, Skype, etc. This fact has motivated a notable increase in the video-traffic supported by cellular networks. In this line, forecasts predict that this type of traffic will represent the 78% of the overall data-traffic supported by mobile networks by 2021 [1]. For that reason, video-streaming techniques should be highly efficient in terms of their network resource demands at the time that they ought to provide satisfactory levels of quality to end-users. An important point regarding the last mile of the end-to-end communication path is the heterogeneity and variability of the elements composing it, e.g., different available radio-transmission technologies, state of the access network, end-user equipment, etc. Therefore, it is necessary that video transport algorithms permit their adjustment in real-time to this wide variety of aspects that heavily determine the quality perceived by the end-user (Quality of user Experience, QoE). In addition, from the service and transport providers' perspectives it is also of high interest the development of next-generation video coding and delivery schemes that may adapt their characteristics to the bandwidth and reliability restrictions that wireless networks inherently present [2].

Under this umbrella, a crucial task for current mobile telcos is monitoring and keeping the provided quality at its expected level, in order to guarantee customers' satisfaction. In this context, QoE measurements are highly valued because they provide key information to the different involved stakeholders, e.g., service providers, network managers, or content producers, among others. Thus, both the academia and the industry are focusing their attention and efforts in this field [3]. Since video services are continuously evolving, new methodologies for monitoring the QoE of these next-generation multimedia services are needed.

Particularly, in this work we take into consideration the ITU-T Recommendation P.1203 [4] because it provides means for evaluating the complete audiovisual, i.e., individual audio and video quality and their interactions, of progressive download and adaptive streaming services over reliable transport protocols. Thus, the main aim of this paper is twofold. First, a functional implementation of the ITU-T P.1203 model is introduced and validated by comparing its performance with a subjective test. This implementation allows the evaluation of the QoE in multimedia services delivered over a reliable transport protocol, namely, TCP (Transmission Control Protocol), and using progressive and adaptive streaming techniques. Thereafter, a case study is carried out, focused on a Live Video Streaming (LVS) service over a 4G (Long Term Evolution, LTE) network. In order to discuss the outcomes from different-nature test-benches and transmission configurations, two experiments have been conducted, one in a real deployment and another using emulation techniques. In both test-benches we have compared the Real-Time Messaging Protocol (RTMP) and the Dynamic Adaptive Streaming over HTTP (DASH) video transport protocols. Therefore, the main contributions of this paper are the following: (i) provide a detailed description of the ITU-T Rec. P.1203 from a hands-on perspective and its implementation, (ii) indicate guidelines for measuring the QoE of a LVS service by means of this novel recommendation, (iii) validate our implementation by comparing its performance with a subjective test and another state-of-the-art implementation, and (iv) compare and discuss the QoE attained by an LVS service applying the ITU-T Rec. P.1203 in a 4G network using two-different TCP-based video transport protocols.

The rest of the document is organized as follows. Section II presents a comprehensive review of the related work. Section III explains the ITU-T Rec. P.1203 and its software implementation process. Section IV describes the research methodology of the case study whose results are shown and discussed in Section V. Finally, the paper ends summarizing the most important findings.

II. RELATED WORK

Due to the richness and the inherent complexity of video-based services, a lot of efforts have been devoted during the last decade to develop proper objective QoE estimation models [5],[6]. Because of the different technologies and scenarios that may be involved in this type of services, developing a general model coping with this heterogeneity has been a difficult task; hence, proposals from both the academia and standardization bodies are intended to be accurate for specific coding and transmission technologies in particular environments. In order to narrow the scope of this review, in the following we focus on those works related to the transmission of video-services using a reliable transport protocol (TCP) over cellular networks.

Garcia-Pineda *et al.* presented [7] a holistic QoE estimation model employing different bit-stream data and basic video quality metrics. The authors applied a statistical method based on factor analysis to evaluate the correlation of the variables, which allowed them to estimate the QoE (in terms of MOS) of the delivered video by following both full-reference and non-reference approaches. Specifically, they focused on the LVS service distributed over a mobile LTE network. In the same context of LTE systems, the authors of [8] and [9] presented two basic studies regarding the impact of different QoS parameters on the final attained QoE of multimedia services. In both cases, the focus was on real-time video services distributed over an LTE infrastructure. Similarly, in [10] the authors studied the effect of several QoS parameters on the QoE of video streaming services using MPEG-DASH, RTSP (Real Time Streaming Protocol), and RTMP (Real Time Messaging Protocol) in an emulated LTE network. They employed the ITU-T Rec. P1201 to assess the QoE level and conclude that new amendments were needed in parametric QoE models to reflect better the behavior of TCP-based video transport protocols. Additionally in [9], a discussion regarding the interaction of QoS/QoE parameters on future 5G and M2M networks was given. In turn, work in [11] proposed a QoE estimation model for progressive video streaming services over mobile broadcast networks. Specifically, the authors proposed a novel methodology for estimating the QoE that takes into account the stalling events suffered by the receiver when visualizing the video stream. Results were validated comparing the attained estimations with the results extracted from subjective tests. Works in [12–14] considered the current buffer occupation in the client side as a key factor for selecting the most proper video rate to be requested for the next segments. Additionally, authors of [14] presented a proposal for improving the QoE in adaptive HTTP video transmissions in cellular environments. Concretely, authors took into consideration both the characteristics of the video content and the available radio resources for overwriting the HTTP requests from the client in a proxy placed within the 4G distribution network. Thus, the data rate of the video segments requested to the content provider was adapted according to the Radio Access Network (RAN) conditions.

In [15], a novel test-bench for evaluating the QoE of 3D video services delivered through an LTE network was developed. The proposed system consisted of three elements: a transmission server, a mobile network emulator based on the NetEm tool, and a video client. The main contribution of this work was the adaptation of the NetEm source code to model the impact of the video packet interarrival delay on the overall attained latency, which directly impacts on the perceived QoE. The authors of [16] presented a machine learning-based methodology for developing a new quality metric using 9 different algorithms. In order to obtain high accuracy in its predictions, the model was trained with a data-set composed of 960 videos sequences, generated under a notable number of conditions that emulated realistic transmissions. A related approach was followed in [17,18]. In the former, the authors used artificial neural networks to develop a QoE estimation model for the IPTV service. In the latter, different QoE models based on different automatic supervised machine learning techniques were employed for predicting the QoE of typical end-user multimedia services such as Youtube and Facebook. As inputs for these models, authors proposed to use passive QoS parameters extracted from the users' terminals.

Very few works have been found in the related literature addressing objective QoE parametric estimation for video streaming services over the TCP protocol, which is the focus of the ITU-T Rec. P.1203. In [19], the authors analyzed the quality prediction performance of the recently standardized parametric P.1203 models for real-streaming in Over-The-Top (OTT) services (YouTube, Vimeo, amazon Instant Video, and a proprietary DASH-based streaming framework). In particular, a validation database comprising bitstream traces from the aforementioned services was used to evaluate the performance of P.1203 (mode 0 and mode 1) models. As a conclusion, the authors proposed that this recommendation could be further enhanced using short-term Full Reference (FR) video-only models. However, the authors did not provide any information about the way the standard P.1203 was applied. Gomez *et al.* presented in [20] a novel LTE mobile wireless testbed for multimedia streaming experimentation. The design of the testbed was based in software-radio, where the typical LTE network architecture (EPC and eNodeB) was provided by a commercial solution, namely, the LTE100 from Amarisoft [21]. The experiments of the case study were focused on events of traffic congestion for evaluating the attained QoE under adverse network conditions. The authors analyzed, in an LTE system, how the downlink bandwidth congestion affected two QoE metrics (the average time a user waits to start watching the video and the average number of re-buffers). In [22], the authors presented a scalable-video quality model, proposing an integral quality prediction for long media sessions for HTTP Adaptive Streaming (HAS). The proposed model was available in four modes of operation for different levels of media-related bitstream information, reflecting different types of encoding of the media stream. This model was chosen by the ITU-T for implementing the video quality estimation module Pv, referenced as the ITU-T Rec. P.1203.1 [23]. Robitza *et al* in [24] described a quality model for HTTP Adaptive Streaming. By considering the recency effect [25] as well as the location and length of buffering events at the player side, this model integrated the different audio and video quality values calculated during the stream-playback into a final quality estimation. It is convenient to clarify that the works [22] and [24] were submitted to the ITU-T P.NATS competition and parts of them have been released in the official recommendation ITU-T P.1203 [4]. Robitza *et al* provided in [26] a set of open data and an implementation of the ITU-T Rec. P.1203. The paper also reported the attained performance in comparison with subjective tests. Authors concluded that significative improvements are obtained when using bit-stream-based video QoE estimation algorithms instead of those based on simple meta-data inspection. They also stressed the robustness achieved when combining classic models with novel approaches based on machine learning for accurately estimating the user's QoE. However, the presented implementation does not explicit the methodology employed for obtaining the initial pre-buffering time as well as the number of stalling events occurred during the video playback. Finally, authors of [27] stated, regarding ITU-T P.1203 model, that its complex definition does not permit a straightforward identification of the underlying assumptions. To overcome this issue, their work investigated the impact of the well-known QoE factors of HAS, namely, initial pre-buffering, stalling events, and rate adaptation, on the QoE output score of the model.

Compared to the state of the art, this work presents a detailed description of the development, implementation, and application of the ITU-T Rec. P.1203. In particular, firstly we provide deep description and discussion of the model and validate our implementation with subjective tests and by comparing it with the implementation presented in [26]. Thereafter, we take as a case study the LVS service considering two TCP-based transport protocols delivered over two different real LTE-based test-benches. Therefore, we aim to provide with a foundational and practical work for the design and development of QoE estimation tools for video-services based on the novel ITU-T Rec. P-1203.

III. A PRACTICAL VIEW OF THE ITU-T REC. P.1203 MODEL

A. *Understanding P.1203*

The ITU-T Rec. P.1203 [4] provides the introductory document for a set of additional ITU-T recommendations that describe the algorithms for monitoring the integral media session quality of TCP-based video streaming services. That is, this recommendation develops a parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. Depending on the number and type of the inspected parameters and the complexity of the processing algorithm, the ITU-T Rec. P.1203 presents four different modes of operation as follows:

- Mode 0: The information is obtained from meta-information available during progressive download or adaptive streaming regarding codec and bitrate, initial loading delay, and stalling events. As an example, the meta-information may be extracted from the manifest files used in DASH.

- Mode 1: All the information from Mode 0 is considered, with additional video and audio frame information based on packet header inspection.

- Mode 2: Starting from the information considered in Mode 1 and up to 2% (in Bytes) of the overall media stream information based on deep packet inspection and partial bitstream parsing.

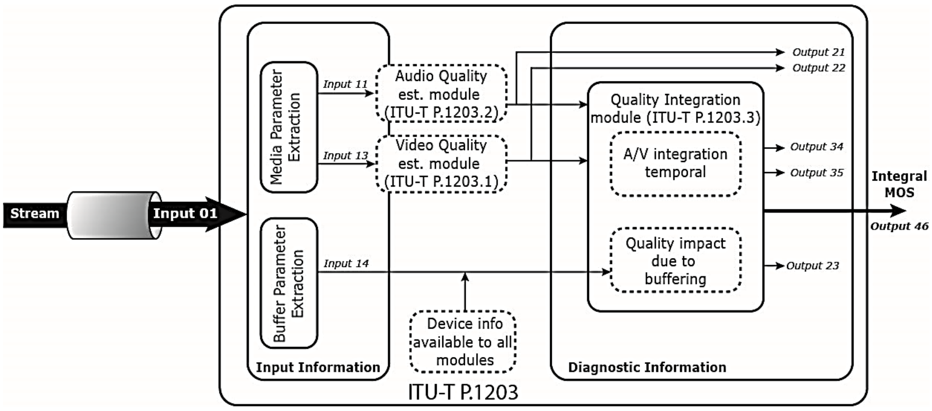

Fig. 1 Modules of the ITU-T P.1203 [4] (I≡Input; O≡Output)

– Mode 3: All the information employed in Mode 1 and the complete media stream information based on bitstream parsing.

The ITU-T Rec. P.1203 model predicts the QoE in terms of MOS [28] on a 5-point absolute category rating (ACR) scale [29]. In addition, the well-defined independent modules provide several diagnostic outputs, which are described in [4]. Fig. 1 shows the different modules of the ITU-T P.1203 model.

As it is shown in Fig. 1, the P.1203 model consists of an audio-quality estimation module Pa [30], a video-quality estimation module Pv [23], and an audio-visual quality integration module Pq [31] gathering the previous ones. Pa and Pv operate on the bitstream-level in a sliding window fashion, where the length of the sliding window is fixed to 20 s. Next, a brief description of each input I, output O, and module that compose the system is provided:

- Stream I.01: Corresponds to the video flow that will be analyzed.
- Media Parameter Extraction Module (MPEM): It is responsible for extracting, depending on the mode of operation, the information and the a priori parameters of the incoming multimedia stream. This module generates the inputs I.11 and I.13.
- I.11: Corresponds to the audio encoding information. The considered parameters are: target audio bit-rate, segment duration, audio frame number, audio frame size, audio frame duration, audio codec, audio sampling frequency, number of audio channels, and audio bit-stream.
- I.13: Corresponds to the video encoding information. The inspected parameters are: target video bit-rate, video frame-rate, segment duration, video encoding resolution, video codec and profile, video frame number, video frame duration, frame presentation timestamp, frame decoding timestamp, video frame size, type of each picture, and video bit-stream.
- Buffer Parameter Extraction Module (BPEM): It extracts the information from the stalling events. In the ITU-T Rec. P.1203 only the following transition states are considered: initial stalling to playing, playing to stalling, playing to end, and stalling to playing. This module generates the input I.14.
- I.14: It corresponds to the stalling parameters employed by the estimation module in charge of evaluating the impact of buffering events. The considered parameters are: stalling/initial loading event start and event duration.
- I.GEN: In this module, the information related to the display resolution and device type is added. The device type is defined as follows: PC/TV (screen size 24 inches or larger and smaller than or equal to 100 inches) and mobile (screen size 10 inches or smaller). It is important to observe that the MPEM, BPEM, and I.GEN modules are not described by the ITU-T Recommendation. Therefore, in order to implement the model it will be necessary to define their functionality for extracting the parameters of interest.
- Pa: Audio Quality Estimation Module: The ITU-T Rec. P.1203.2, part of ITU-T Rec. P.1203, details the module that implements the bitstream-based short-term audio quality estimation. The algorithm proposed for implementing this model was presented in [30], and it is defined in (1) (2) and (3), with Bitrate being the audio bit rate in kbit/s.

$$O.21 = MOSfromR(QA) \tag{1}$$

$$QA = 100 - QcodA \tag{2}$$

$$QcodA = a1A * e^{(a2A * Bitrate)} + a3A \tag{3}$$

The function $MOSfromR$ is included in the Annex E of [23] and is obtained as follows:

$$MOSfromR: \mathbb{R} \longmapsto \mathbb{R}$$

$$Q \longmapsto MOS := MOSfromR(Q)$$

$$MOS = MOS_{MIN} + (MOS_{MAX} - MOS_{MIN}) * \frac{Q}{100} + Q * (Q - 60) * (100 - Q) * 0.000007 \tag{4}$$

$$MOS = min\big(MOS_{MAX}, max(MOS, MOS_{MIN})\big) \tag{5}$$

where $MOS_{MAX} = 4.9$ and $MOS_{MIN} = 1.05$. Coefficients *a1A*, *a2A*, and *a3A* depend on the audio codec; these are described in the Table 8-1 in [30]. MOS in (5) corresponds to the output O.21 (audio coding quality per output sampling interval), which provides per-one-second scores per session in a 1-5 quality scale.

- Pv: Video Quality Estimation Module (ITU-T P.1203.1): This block estimates the video quality of the received stream. The algorithm recommended by the standard for its implementation is described in [23]. This module finds the overall degradation suffered by the incoming stream due to video representation (spatial and temporal sampling and video compression), which is quantified by the D parameter. D is calculated based on the quantization (*Dq*), upscaling (*Du*), and temporal (*Dt*) degradations as shown in (6)-(14).

$$D = max(min(Dq + Du + Dt, 100), 0) \tag{6}$$

$$Dq = 100 - RfromMOS(\widehat{MOS}q) \tag{7}$$

$$Du = u1 \cdot log_{10}(u2 \cdot (scaleFactor - 1) + 1)) \tag{8}$$

$$Dt = \begin{cases} D_{\mathrm{t1}} - D_{\mathrm{t2}} - D_{t3}, \mathrm{framerate} < 24 \\ 0, \mathrm{framerate} \geq 24 \end{cases} \tag{9}$$

$$D_{t1} = \frac{100 \cdot (t1 - t2 \cdot framerate)}{t3 + framerate} \tag{10}$$

$$D_{t2} = \frac{\mathrm{Dq} \cdot (\mathrm{t1} - \mathrm{t2} \cdot \mathrm{framerate})}{\mathrm{t3} + \mathrm{framerate}} \tag{11}$$

$$D_{t3} = \frac{\mathrm{Du} \cdot (\mathrm{t1} - \mathrm{t2} \cdot \mathrm{framerate})}{\mathrm{t3} + \mathrm{framerate}} \tag{12}$$

$$D_t = max\ (min(Dt, 100), 0) \tag{13}$$

The Maximum quality according to the pure upscaling and temporal degradations is calculated as follows:

$$Qmax = \begin{cases} 100 - \widehat{\mathrm{Du} - \mathrm{D}}\mathrm{t1}, \mathrm{framerate} < 24 \\ 100 - Du, framerate \geq 24 \end{cases} \tag{14}$$

Then, quality (Q) is calculated as follows:

$$\hat{\mathrm{Q}} = 100 - \mathrm{D} \tag{15}$$

Finally, (16) gets the value of video quality estimation, which corresponds to the output of the Pv module.

$$\widehat{MOS} = MOSfromR\big(\hat{Q}\big) = O.22 \tag{16}$$

Where $\hat{Q}$ is the estimated video encoding, with $\hat{Q} \in [0,100]$. $MOSfromR$ is described in Annex E of [23] and D is the overall degradation due to video representation (spatial and temporal sampling, video compression).

- Pq: Quality Integration Module (ITU-T P.1203.3) [31]: This module predicts the impact of audiovisual quality variations and stalling events on quality experienced by the end-user in multimedia mobile streaming and fixed network applications using adaptive bit-rate streaming. This module must receive information about the estimated quality of the audio (O.21) and video (O.22) streams and the occurrence of stalling events during playback (I.14, I.GEN). It is composed of two sub-modules: Pav: A/V and Pb.
- Pav: A/V Integration/temporal: The outputs of this module are the audiovisual segment coding quality per output sampling interval (O.34) and the final audiovisual coding quality score (O.35). O.34 is derived from O.21 and O.22 as shown in (17), where $t = [1,2,\ldots,T]$, *av1*=–0.00069084, *av2*=0.15374283, *av3*= 0.97153861, and *av4*= 0.02461776.

$$O.34_t = max(min(av_1 + av_2 * O.21_t + av_3 * O.22_t + av_4 * O.21_t * O.22_t, 5),1) \quad (17)$$

The final audiovisual coding quality (O.35) considers the audiovisual quality per output sampling interval (O.34) as well as any temporal effects and media length. The calculation of O.35 is described in section 8.3 of [31].

- Pb: Quality Impact due to buffering: The output of this module is the perceptual stalling indication O.23 (18). It is calculated based on the number of stalling events (*numStalls*), the weighted sum of stalling events (*totalStallLen*), and the average of the time interval between stalling events (*avgStallInterval*), where s_1=9.35158684, s_2=0.91890815, and s_3=11.056755.

$$O.23 = 1 + 4 * \mathrm{SI} \quad (18)$$

$$SI = exp^{(-numStalls/s_1)} * exp^{\left(-\frac{\mathrm{totalStallLen}}{T}/s_2\right)} * exp^{\left(-\frac{\mathrm{avgStallInterval}}{T}/s_3\right)} \quad (19)$$

- Model output O.46: This final media session quality score is calculated based on the stalling events, the final audiovisual compression quality (O.35), the Random Forest model prediction (*RFPrediction*), and media length. Initially, a temporary value of O.46 is calculated (20) and then is compensated by the differences in subjective ratings due to the heterogeneity of tests across different laboratories (21). For the calculation of *RFPrediction*, the machine learning module for audiovisual quality estimation introduces the features (with their IDs and values) described in clause 8.1.3 of [31] in an ensemble of 20 decision trees, namely, a Random Forest. Each decision tree has a maximum depth of 6.

$$O.46\_temp = 0.75 * (1 + (O.35 - 1) * SI) + 0.25 * RFPrediction \quad (20)$$

$$O.46 = 0.02833052 + 0.98117059 \cdot O.46\_temp \quad (21)$$

It is important to observe that in order to use the quality integration module Pq, i.e., the ITU-T P.1203.3, it is necessary to comply with certain factors and applications ranges summarized in Table 1.

Algorithm 1 MOS estimation algorithm for different scenarios (Real, Hybrid, Naka0.5, Naka1, Naka5 & Naka7) and protocols (480-RTMP, 720-RTMP, 480-DASH & 720-DASH)

Input: Constants and Stalling events
Output: $MOS_{estimation}$
1: **Constant definition**
2: **Read Stalling events parameters**
    $Events_{time\text{-}stamp}$, $Events_{length}$, $Events_{number}$,$Playback_{length}$
3: **O21 & 022 estimation**
    According to:
        3.1. Audio & Video bit-rate
        3.2. Emission & Reception codification
4: **$RF_{parameters}$**
    for *each scenario* **do**
        for *each protocol* **do**
            for *each test* **do**
                avgBuffInterval, totalBuffLen, O34, $RF_{parameters(1\text{-}14)}$ estimation
            end *test* **for**
        end *protocol* **for**
    end *scenarios* **for**
5: **$RF_{prediction}$**
    for *each scenario* **do**
        for *each protocol* **do**
            for *each test* **do**
                Obtain $RF_{prediction\ per\text{-}tree}$ value using RandomForest
            end *test* **for**
            $RF_{prediction}$ = ***mean(*** *$RF_{prediction\ per\text{-}tree}$)*
        end *protocol* **for**
    end *scenarios* **for**
6: **$MOS_{estimation}$**
    for *each scenario* **do**
        for *each protocol* **do**
            for *each test* **do**
                O35, O23, vidQualChangeRate, and O46 estimation
            end *test* **for**
            $MOS_{estimation}$ = ***mean(*** *$O46_{per\text{-}test}$)*
        end *protocol* **for**
    end *scenarios* **for**
7: **Result Representation**

Algorithm 1. Our ITU-T P.1203 model implementation.

TABLE 1. FACTORS AND APPLICATION RANGES OF THE ITU-T P.1203.3

Video sequence duration	60 seconds – 5 minutes
Initial loading delay and stalling	0-10 seconds
Maximum number of stalling events	5
Maximum length of a single stalling event	15 seconds
Total stalling duration	30 seconds
Other details	No stalling within 5 seconds of the start of the video playing.

B. Software implementation of the ITU-T P.1203 model

Following the ITU-T Recommendation under study, we have developed a functional implementation of the P.1203 model with the aim of evaluating the QoE in terms of MOS of a live video streaming service over LTE. Thus, the developed algorithm computes the estimated MOS value of a given session according to the number and duration of stalling events as well as the total playback-related and other parameters defined by the recommendation. Algorithm 1 includes the implemented process for all tests under study. As aforementioned, we will carry out experiments in two different test-benches, namely, a real scenario and an emulated one, and then two protocols are compared, RTMP and DASH, using different resolutions. RTMP is a standard for streaming multimedia content from a server to a remote media player over Internet using TCP. DASH is an adaptive streaming standard where a client requests multimedia content using HTTP to standard web servers. The values 480 and 720 indicate the minimum number of pixels in transmitted or received video in height (H) or width (W).

As observed in Algorithm 1, first, we declare all the constant variables defined on the Recommendation, which are used to compute both final and intermediate parameters. In this line, the ITU-T Rec. P.1203.1 [23] defines all constants and variables for video quality estimation, the ITU-T Rec. P.1203.2 [30] focuses on those for audio quality estimation, and finally, the ITU-T Rec. P.1203.3 [31] includes multimedia and final score integration resources for sessions between 30s and 5 min.

Then, the algorithm loads the stalling events parameters such as *event time-stamp*, which indicates when a given event starts, *event length*, that indicates how long the event is, *number of events* during the playback, and *total playback length*. Along the code, these parameters are used several times because they are key to estimate other final or intermediate variables. It is important to observe that stalling events are key performance metrics when measuring the QoE. As an example, several recent works from the related literature focus on them to improve QoE. In [3], the authors proposed to apply deep Q-learning improvements to the neural network architecture and the learning mechanism to decrease the number and duration of stalling events to a minimum while maintaining a high video bit rate. Ghadiyaram *et al*. propose in [32] a QoE evaluator, called QoE Indexer, which takes into account the interactions of stalling events in the measurement of QoE. The predictions are made in continuous time and the results obtained are approximate to subjective QoE measurements. Additionally, the same authors presented in [33] a database of 174 videos affected by distortions caused by 26 different patterns of stalling events. They showed the methodology and results of the subjective evaluation of QoE on 54 individuals.

In phase 3 of Algorithm 1, the processing blocks for estimating the different types of quality estimations come to play. The O21 parameter is calculated, which is the audio quality per output sampling interval on the MOS scale, as well as the O22 output, which is the equivalent for video quality using also the MOS scale. Depending on the video-transport protocol under study (480-RTMP, 720-RTMP, 480-DASH and 720-DASH) different constant are used: *audio bit-rate* is the required memory space to store one second of audio, *video bit-rate* is the information that is played per second, *emission codification*, which converts the information to send it to the receiver, *reception codification*, which translates the received message and adapts it to be played, and finally *framerate*, which is the number of frames per second in the video.

The next step is used to calculate a set of intermediate variables using the information from the stalling events and the constants defined in the ITU-T Rec. P.1203. These variables are obtained for each test, protocol, and scenario. In particular, the variables are: *avgBuffInterval*, which is the average interval between stalling events, *totalBuffLen*, which is the total length of stalling events taking into account the buffer weight, O34, which is the audiovisual quality per output sampling interval derived from O21 and O22, and 14 $RF_{parameters}$, which are described in Table 2.

Once we have the $RF_{parameters}$ and after applying the ITU-T P.1203.3 model, the $RF_{prediction}$ value is extracted. $RF_{prediction}$ is obtained from a particular supervised machine learning algorithm, called Random Forest. The RF model included in the recommendation has 20 trees, where each tree has a maximum depth of 6. Each row of the tree has 5 pieces of information: *nodeID*, *featureID*, feature threshold, left child *nodeID*, and right child *nodeID*. Each decision tree starts at node 0 comparing the *featureID* ($\mathbb{Z} \in [1,14]$) with the analogous

TABLE 2. PARAMETERS RELATED TO THE MACHINE LEARNING MODULE

Feature	Feature Name	Description
0	*stallCountWithoutInitial*	Total number of stalling events occurring in the media session, excluding the initial stalling event.
1	*stallDur*	This is the sum of duration of all the stalling events.
2	*stallFreq*	Number of stalling events dividing by the length of media.
3	*stallRatio*	Ratio of stalling events is ratio of stallDur to the total media length.
4	*timeLastStallToEnd*	Time elapsed since the start of occurrence of last stalling event to the end of video.
5	*averagePvScoreOne*	Average of all the *O.22* scores that correspond to the first third of the *O.22* scores vector.
6	*averagePvScoreTwo*	Average of all the *O.22* scores of the second third of the *O.22* scores vector.
7	*averagePvScoreThree*	Average of all the *O.22* scores of the third third of the *O.22* scores vector.
8	*1PercentilePvScore*	1st percentile of *O.22*.
9	*5PercentilePvScore*	5th percentile of *O.22*.
10	*10PercentilePvScore*	10th percentile of *O.22*.
11	*averagePaScoreOne*	Average of *O21* scores in the first half of *O21* score vector..
12	*averagePaScoreTwo*	Average of *O21* scores in the second half of *O21* score vector
13	*T*	Length of the media.

$RF_{parameter}$. If the $RF_{parameter}$ is below the feature threshold on the corresponding node, it moves to the node with the left child *nodeID*. Conversely, if $RF_{parameter}$ is above the feature threshold, it moves to the node with right child *nodeID*. The process continues until reaching a *featureID* value of *-1*, which indicates that the actual node is a leaf node and it is reached. The feature threshold value of a leaf node is the $RF_{prediction}$ to this particular decision tree. The final value of $RF_{prediction}$ for a test is the mean of the output of the 20 trees.

Lastly, the $MOS_{estimation}$ process uses the stalling event parameters, the pre-provided constants, and the $RF_{prediction}$ value obtained before to assess the final quality score. Some intermediate values are also necessary as explained earlier: O35, which is the final audiovisual coding quality, O23, which is a perceptual stalling indicator, *vidQualChangeRate* that is the difference between the maximum and minimum video quality score within O22, and $O46_{per\text{-}test}$, which is the final media session quality score obtained per test. The final $MOS_{estimation}$ value is calculated for each protocol in each scenario as the mean of all $O46_{per\text{-}test}$ values for all tests.

IV. METHODOLOGY

Once we have explained the ITU-T Rec. P.1203 model and its implementation, we validate it and put it in practice using a complete case study. In the following, we detail the tools employed in our experiments and describe the methodologies adopted in both the validation and application phases.

A. Test-benches

A.1. Real Test-bench

The core element of the deployed architecture was the CMW500 communication analyzer by Rohde & Schwarz. This equipment supports different wireless communication technologies such as LTE-A, WCDMA/HSPA+, GSM/GPRS/EGPRS, CDMA2000 or TD-SCDMA, among others. For this study, we just focused on the LTE-A technology whose installed version complied with the 3GPP's 36.521-1 standard [34]. The CMW500 is able to operate as an 4G eNodeB as well as receiving a source of data (video in our case), which will be accessible through the deployed LTE network. Therefore, a video server (LVS server in Fig. 2) was connected to the CMW500 through its Data Application Unit (DAU). The video client, a regular Windows-based laptop, was placed at a distance of 50 m apart from the CMW500 and connected to the LTE network by means of a 4G ZTEMF823 USB modem. Finally, the video client was able to connect and consume video content from the LVS server in order to perform the quality-metrics measurements.

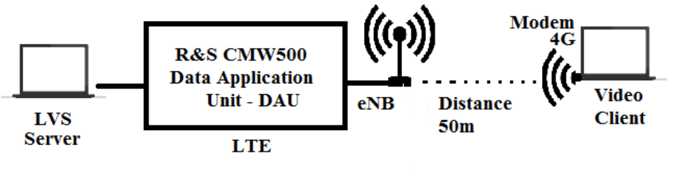

Fig. 2 Real test-bench

A.2. Emulation test-bench

Fig. 3 shows the elements composing this scenario. This network topology was taken from that proposed in [35] and validated by the authors in [36] and [37]. In this case, our deployment was formed by three computers. PC2 and PC3 acted as video server and client, respectively. PC1 hosted the simulated LTE system by means of the NS-3.26. This module simulates the complete LTE stack and allows the deployment of an LTE network composed of a series of eNodesB and terminals. This framework was installed in a Linux Ubuntu 12.04 machine with an Intel Core 2 processor (2.13 GHz), with 4 GB of RAM. PC2 and PC3 were connected to PC1 through their ethernet interfaces. PC2 used Windows 7 Professional with an Intel Core i7-3612QM processor (2.1GHz) and 8 GB of RAM. Wowza Streaming Engine [38] was installed on PC2 to serve the video contents that were consumed by PC3. PC3 also used Windows 7 Professional with an Intel Core 2 Duo processor (2.1GHz x2), 4 GB of RAM, and a video-client based on HTTP Apache. In order to obtain the quality metrics needed for feeding the QoE model considered in this work, the Wireshark protocol analyzer [39] was installed in both communication end-points, i.e., PC2 and PC3; hence, the real traces of the video transmission from the video server and client could be extracted for further analysis.

Regarding the simulation of the propagation conditions in the simulated LTE system, we chose the Nakagami-*m* [40] propagation model, as it is one of the most precise probabilistic models for characterizing outdoor propagation conditions. This model is defined by the shaping factor *m* that characterizes the environmental fading level. Thus, when m=1, the Nakagami-*m* Probability Density Function (PDF) tends to the Rayleigh PDF. With higher values of *m*, the Nakagami-*m* PDF tends to the Rician one. Therefore, it can be seen that the fading level tends to 0 when *m* tends to ∞. Due to its versatility, the Nakagami-*m* model is appropriate for describing the propagation conditions of diverse complex scenarios such as the urban ones. In addition, for comparison purposes, the HybridBuildings propagation model was also employed. HybridBuildings results from the integration of different propagation models and provides propagation-loss estimations in the frequency range between 200 MHz and 2600 MHz for different environments and conditions (indoor and outdoor). This model includes the predictions provided by the Hata, COST231, and ITU-T Rec. P.1411 and P.1238 [41].

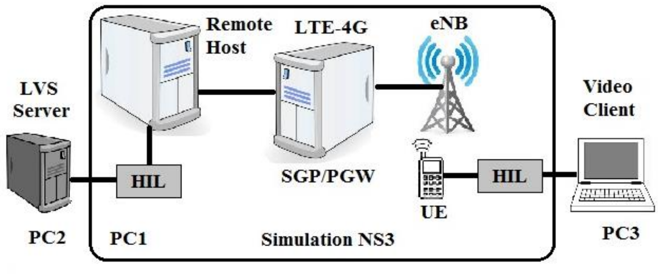

Fig. 3 Emulation test-bench [36].

C. Common features to both test-benches

RTMP and DASH, two widely-used video-transport protocols, were used in the two test-benches. Both protocols employ TCP as a reliable transport-layer protocol. In all the conducted experiments, we used a H.264 sequence of 180 s extracted from the well-known "Big Buck Bunny" short cartoon-movie. Each experiment, considering specific conditions such as resolution (480p and 720p), video-transport protocol (RTMP and DASH), propagation model (Nakagami-m with m equal to 0.5, 1, 3, 5, 7 and HybridBuildings), and type of test-bench (real and emulation), was repeated 10 times for avoiding unexpected singularities. Therefore, we carried out a total of 280 tests, gathering the corresponding data sets for each of them. Finally, by extracting the needed metrics from the collected traces, including packet loss, latency, stalling events, etc., the ITU-T P.1203 model was implemented in Matlab in order to obtain the QoE in terms of MOS for each video transmission. Table 3 and Table 4 present the technical configuration parameters for each of the considered test-benches.

A.4 Information from stalling events

One of the most important challenges to apply the ITU-T Rec. P.1203 is how to obtain the information from the stalling events [32], [33]. Therefore, we consider of interest to briefly explain this process in our case study. For each video transmission, a stalling event file was created with the structure shown in Table 5. Each row indicates the beginning time of one event and its duration. Particularly for this example, there are 6.17 s of initial pre-buffering, a first stalling event of 5.42 s after 85.39 s of playback, and a second stalling event lasting 3.12 s that started in the playback time equal to 137.18 s. Finally, the total playback time is also shown. Next, from all collected stalling event files, the 14 features ($RF_{parameters}$) are obtained to compute the $RF_{prediction}$ factor from the Random Forest model. As an example, Table 6 shows the $RF_{parameters}$ and the $RF_{prediction}$ calculated for the 10 repetitions of one of the experiments.

TABLE 3. CONFIGURATION OF THE EMULATION TEST-BENCH

Parameter	Description Test-Bench	
	Emulation	Real
eNB Operation Power	100dBm	RSRP 51 (-90 to -89 dBm)
Physical Layer Profile	OFDMA, FDD, Band 7, Frec. DL= 2644 MHz, Frec. UL=2535, BW cell UL=DL=10 MHz, Nro RB: DL=UL=50. Mod. QPSK	
Antenna type Path Loss Model	Cosine, height 1.5 m *Nakagami* (*m*= 0.5; 1; 3; 5; 7) *and HybridBuildings*	Directional antenna panel 14 IP-G14-F2425-HV, height 2.5 m
UE Operation Power	26 dBm, height 2.5m, distance 50m	height 2.5m, distance 50m
eNB Noise Figure UE Noise Figure	2 dB 7 dB	
Transmission Mode	SISO	

TABLE 4. CODING SETTINGS

Parameter	Description
Video encoding	
Encoder	H.264/MPEG-4 AVC
Resolution (px)	480p (854x480) and 720p (1280x720).
Coding bit-rate (Kbps)	1128 for 480p and 2628 for 720p
Frames per second	30
Audio encoding	
Codificador	AAC
Channels	Stereo
Bitrate (Kbps)	128 for 480p and 192 for 720p
Frequency sample	44.100 z

TABLE 5. EXAMPLE OF A STALLING EVENTS FILE

Event Time	Event Duration	Tplayback	Description
0.00	6.17		Initial pre-buffering
85.39	5.42		Stalling event
137.18	3.12		Stalling event
		180	Total playback

B. Use case

For the use case, we measure the QoE attained in a typical LVS service delivered over an LTE network in the two test-benches described above. Several parameters have been considered in both scenarios: video transmission protocol (RTMP and DASH), resolution (480p and 720p), and propagation models in outdoor environments (Nakagami-*m* and HybridBuildings). Comparing the emulated scenario with the real one in terms of QoE, we can, for example, determine the best propagation model in outdoor environments to be used in emulation, i.e., the propagation model that better matches a real LTE deployment. That will be the functionality of our case study, as an example of the wide range of possibilities that the use of a standardized QoE measurement model as the P.1203 can provide.

C. Validation test

As aforementioned, we have conducted a subjective test in the University of Quindío (Colombia) labs, following the guidelines in ITU-T P.910 [29] and ITU-T P.913 [42]. In this test, 40 individuals participated by evaluating the QoE (in terms of MOS) perceived when consuming a series of video sequences characterized as explained in the following. The original video sequence was a 100 s extract of the "Big Buck Bunny" short cartoon-movie (the same as in the use-case). It was codified at 720p with the codec features shown in Table 4 and encapsulated and transmitted using DASH. The network conditions were emulated employing the emulation test-bench and with the configuration showed in Table 3. The considered propagation model was Nakagami-m with m=5 because, as showed later, it is the configuration that better matches real outdoor transmission conditions. The UE was placed 30 m apart from the base station and, additionally, we have introduced additional controlled delays of {0, 25, 50, 75, 100} ms to the transmissions in order to evaluate the system under different network conditions. For each scenario, we repeated the QoE evaluation for 10 times, so we present the average MOS of the user panel for each transmission conditions.

V. Results

In this Section, we first present the validation results of our implementation and, then, the outcomes extracted from the use case under consideration are reported and discussed.

A. Validation test

Fig. 4 presents the QoE estimation (in terms of MOS) obtained in (i) the subjective test, (ii) by our implementation, and (iii) by the implementation taken for comparison purposes [26], when introducing controlled delays in the communication path. It can be observed how our implementation correlates better than the other proposal with the outcomes from the subjective test. In addition, it also evidenced that our implementation is more reactive to the effect of latency variations in the network. The results of the other proposal are very similar in two well-distinguished blocks: from 0 ms to 50 ms, and from 75 ms to 100 ms. In turn, our solution shows a smooth MOS decrease when the delay is increased. Besides, observe that in most of the cases, the confidence intervals are also shorter in the results attained by our implementation. Overall, it can be seen that our proposal is more accurate in relation with the subjective test than the proposal in [26].

In order to confirm these results from a statistical perspective, we have conducted an Analysis of Variance (ANOVA) test with the following null hypothesis: "There are not significative statistical differences between the QoE estimations obtained in the subjective test and those attained by the ITU-T P.1203 implementations". Table 6 shows the square average $\overline{x^2y^2}$; the F factor, the P probability, the critical factor F_c, and the Pearson coefficient r_1. As for both implementations $F \leq F_c$, it can be confirmed the null hypothesis is true, i.e., both solutions correlate well with the subjective test. Focusing on P, it can be concluded that the implementation that provides closer results to those of the subjective test is ours, as the attained P is greater than that obtained by the alternative model. Besides, by analyzing r_1, it is noticed that our proposal has a greater positive linear correlation with the subjective methodology. In the light of these outcomes, it can be concluded that the tool presented in this work provides QoE estimations more correlated with subjective tests than the other state-of-the-art implementation [26], although the latter also provides accurate results.

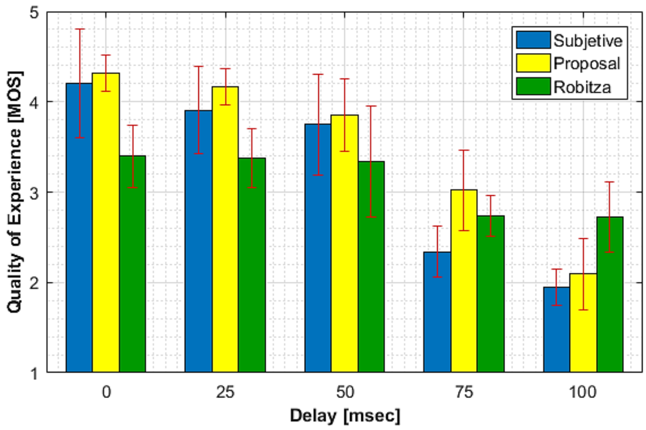

Fig. 4 QoE estimations attained in the validation test.

Table 6. ANOVA and Pearson coefficient obtained in the validation test

Method	$\overline{x^2y^2}$	*F*	*P*	*Fc*	r_1
Proposal and Subjetive	0.03396	0.05936	0.81363	5.31765	0.9912
Robitza and Subjetive	0.09581	0.10017	0.759726	5.3176550	0.9866

Table 7. Example of $RF_{PARAMETERS}$ and $RF_{PREDICTION}$ Factor

Feature	Feature Name	Test1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
0	*stallCountWithoutInitial*	1	1	2	1	1	1	1	2	1	1
1	*stallDur*	1.333	1.333	3.333	3.333	3.333	1.333	1.333	3.333	3.333	3.333
2	*stallFreq*	0.007	0.008	0.014	0.008	0.008	0.007	0.008	0.014	0.008	0.008
3	*stallRatio*	0.009	0.010	0.024	0.025	0.025	0.009	0.010	0.024	0.025	0.025
4	*timeLastStallToEnd*	120	118	55	123	123	120	118	55	123	123
5	*averagePvScoreOne*	3.471	3.944	3.719	3.915	3.915	3.471	3.944	3.719	3.915	3.915
6	*averagePvScoreTwo*	3.471	3.944	3.719	3.915	3.915	3.471	3.944	3.719	3.915	3.915
7	*averagePvScoreThree*	3.471	3.944	3.719	3.915	3.915	3.471	3.944	3.719	3.915	3.915
8	*1PercentilePvScore*	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339
9	*5PercentilePvScore*	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339
10	*10PercentilePvScore*	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339	4.339
11	*averagePaScoreOne*	3.647	4.145	3.908	4.114	4.114	3.647	4.145	3.908	4.114	4.114
12	*averagePaScoreTwo*	4.863	5.526	5.211	5.485	5.485	4.863	5.526	5.211	5.485	5.485
13	*T*	8	7	7	7	7	8	7	7	7	7
RF Prediction Factor		**4.471**	**4.503**	**4.296**	**4.370**	**4.370**	**4.471**	**4.503**	**4.296**	**4.370**	**4.370**

B. Use case

In the following, we present the outcomes extracted from the use case previously described. Fig. 5 and Fig. 6 show the level of QoE in terms of MOS obtained by applying the ITU-T Rec. P.1203 model. These figures include the QoE, that is O.46 as depicted in (21), for the two different test-benches (real and emulated LTE system), employing RTMP and DASH protocols with two different resolutions, 480p and 720p, and the two propagation models under consideration. The bar graphs present the average values and confidence intervals ($\alpha = 0.05$) attained for each configuration, hence avoiding non-representative singularities. From these figures, it can be seen that the propagation model that better fits the behavior of the system in real conditions is the Nakagami-*m* model, with m = 5. Lower values of m imply most adverse propagation conditions such as those of indoor environments, and higher values tend to represent almost free-space conditions, which was not the case of the real environment where the experimental tests were conducted. Regarding the performance of both transport protocols, observe in Fig. 5 how DASH presents higher MOS values than RTMP almost all the studied conditions, especially with the highest resolution. In turn, a MOS decay is evidenced in Fig. 6 when using the highest resolution (720p) for both codecs. This is explained by the greater volume of data transmitted with higher resolutions that provokes transmission degradations in the form of packet loss. As discussed later, the presence of losses leads to a increase on the number and duration of the stalling events, which have a great weight in the QoE estimation of the ITU-T Rec. P.1203 as discussed in the following.

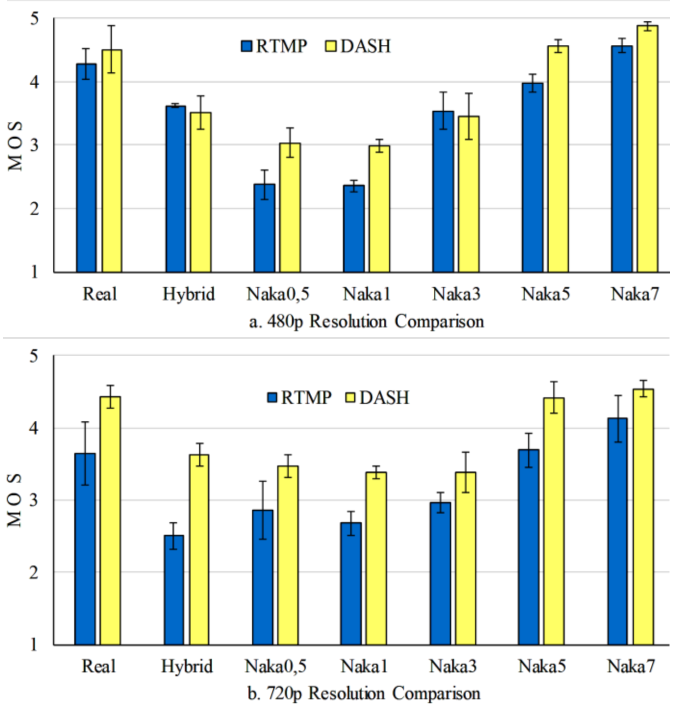

Fig. 5 Average QoE measurements using the ITU-T Rec. P.1203 model (including confidence intervals with α=0.05) using different propagation models, resolutions, and protocols (a. 480 and b. 720).

Fig. 7 presents a comparison of the number and duration of stalling events with the estimated QoE. First, comparing Fig. 7(a) and Fig. 7(c), it is clear that an increase in the number of stalling events leads to a notable decrease in the MOS levels. The same applies to the duration of the stalling events (compare Fig. 7b and Fig. 7c). Second, the number of stalling events is greater for RTMP than for DASH (Fig. 7a), which causes lower levels of MOS for RTMP as discussed previously. On the contrary, the duration of the stalling events is shorter for the RTPM protocol than for DASH (see Fig. 7b). From the above, it is evident a direct relationship between the number and duration of stalling events with the value of MOS; this relationship is also obtained by Seufert *et al* in [27]. The lower the number of stalling events and the shorter, the better MOS is obtained. However, there is also an unnoticeable result, which is that the number of stalling events has more effect on the final QoE than the duration of those events. With a resolution of 480p using the DASH protocol and the Nakagami-*m* (m = 7) propagation model no stalling events were detected, as expected for almost ideal propagation conditions. On the other hand, the greater number of stalling events and longer durations are obtained with the most adverse scenarios, i.e., those characterized with the Nakagami model with m=0,5 and m=1. Fig. 7c also shows the value of two intermediate outcomes: O.21 (audio quality per output sampling interval, Pa) and O.22 (video quality per output sampling interval, Pv). These values are key for the calculation of the final output (O.46, integral MOS) because this outcome is derived by the Quality

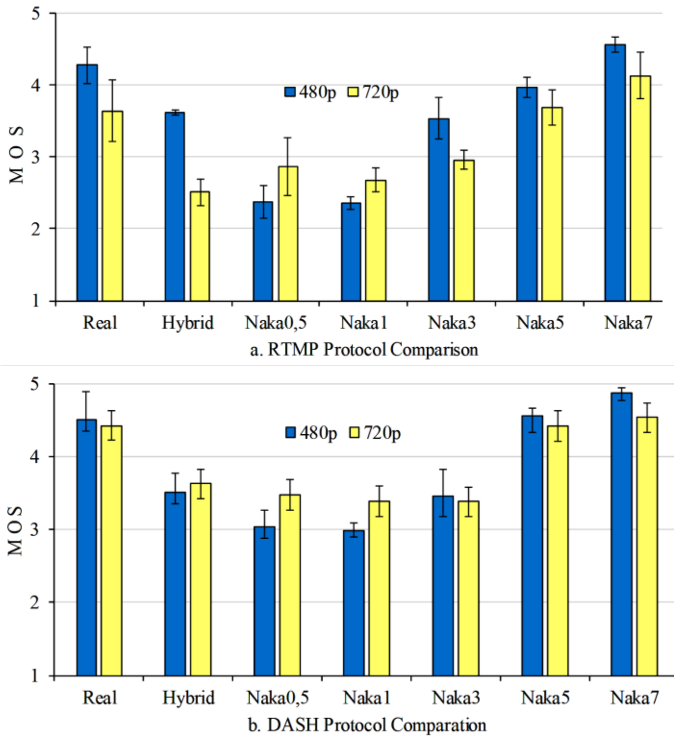

Fig. 6 Average QoE measurements using the ITU-T Rec. P.1203 model (including confidence intervals with α=0.05) using different propagation models, resolutions, and protocols. a. RTMP; b. DASH.

Integration Module together with the information of stalling events (I.14). Other output values such as O.34 (audiovisual segment coding quality per output sampling interval) and O.35 (final audiovisual coding quality score) are derived from O.21, O.22, and I.14 as well. Observe that the values for O.21 and O.22 remained constant all along the experiments as they do not depend on the transmission conditions but only on the video/audio encoding characteristics (included in Table 3). Therefore, considering these preliminary and constant quality parameters, the final estimated MOS is determined by the transmission conditions which are evidenced in the user's terminal and evaluated by the model in the form of stalling events.

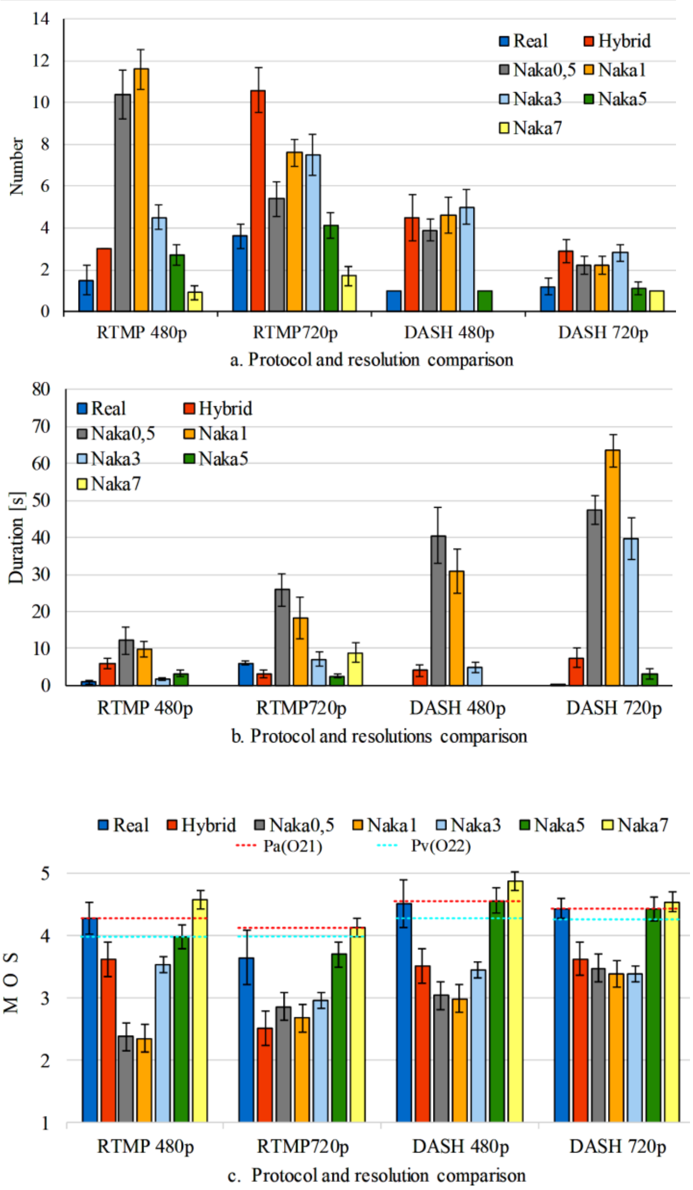

Fig. 7 Comparing the average number of stalling events and their duration (both metrics include confidence intervals α=0.05) with the QoE MOS (O.46), Pa (O.21), and Pv (O.22) obtained from the ITU-T Rec. P.1203 model.

Finally, observe that the propagation model tested in the emulation test-bench that better matches the results obtained in the experimental test-bench (the real scenario) is Nakagami-*m* with $m = 5$. This result corroborates previous outcomes from the same authors regarding the most proper configuration of LTE in emulation tools [36]; where test-benches are validated from QoS parameters, such as: packet loss rate (PLR), jitter, interpacket delay and throughput. From the above, it can be affirmed that the use of the presented platform will allow studies, such as traffic characterization for the LVS service on an emulated LTE network, which offers a high degree of reliability due to the accuracy of the proposed platform with the real scenario and greater flexibility than real deployments.

VI. CONCLUSION

In this work, it has been presented a functional implementation of the ITU-T P.1203 model. This implementation allowed the evaluation in terms of QoE of multimedia services delivered over a reliable transport protocol, namely, TCP, and using progressive and adaptive streaming techniques. We have validated this tool by contrasting its correlation with the results extracted from a subjective test and by comparing its performance with that of another state-of-the-art implementation. We also evaluated a case study focused on an LVS service over an LTE network. We focused as well on validating the most appropriate propagation model included in the simulation environment (NS3) in order to accurately mimic realistic outdoor propagation conditions. From the attained results, several conclusions can be extracted. The DASH protocol presented higher levels of QoE (MOS) in comparison with RTMP. In turn, by analyzing the system performance with the considered resolutions (480p and 720p), in general the estimation of the MOS was greater for the resolution of 480p in comparison with that attained for 720p. We assumed this was caused by the greater number of stalling events suffered when using a higher resolution due to the greater volume of transmitted data. In this line, the relationship between the number and duration of stalling events in the evaluation of the MOS was evident, and the number of stallings events seemed to be more decisive than the duration of those events. Finally, we showed that the Nakagami-*m* propagation model characterized with $m = 5$ is the model that better mimicked the outdoor propagation conditions under consideration. To sum up, by means of the ITU-T Rec. P.1203 implementation presented in this work, it was possible to obtain the objective QoE estimation in a cost-effective and low-complexity manner. As future work, we plan to extend the evaluated scenarios and corroborate our measurements with additional subjective tests.

ACKNOWLEDGMENT

This work has the support of Grupo de Ingeniería Telemática of the Universidad Politécnica de Cartagena (Spain), Grupo de Ingeniería Telemática of the University of Cauca (Colombia), and Grupo de Investigación en Telecomunicaciones of the University of Quindío (Colombia). In addition, it counts with the economic support of

COLCIENCIAS through the call Doctorado-Nacional-647, University of Quindio, and the AEI/FEDER UE project grant TEC2016-76465-C2-1-R (AIM).

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper, Cisco. (2017).

[2] H.G. Msakni, H. Youssef, Ensuring video QoE using HTTP Adaptive Streaming: Issues and challenges, in: 2016 5th Int. Conf. Multimed. Comput. Syst. ICMCS, 2016: pp. 200–205. doi:10.1109/ICMCS.2016.7905586.

[3] J. Liu, X. Tao, J. Lu, QoE-Oriented Rate Adaptation for DASH With Enhanced Deep Q-Learning, IEEE Access. 7 (2019) 8454–8469. doi:10.1109/ACCESS.2018.2889999.

[4] ITU-T Recommendation, P.1203 : Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Video quality estimation module, (2017). http://www.itu.int/rec/T-REC-P.1203.1-201710-I (accessed February 2, 2019).

[5] H.F. Bermúdez, J.L. Arciniegas, E. Astaiza, State of the art of QoE assessment methods, and emulation environments for the video service in LTE networks, Entre Cienc. E Ing. 10 (2016) 66–75. http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S1909-83672016000200010&lng=en&nrm=iso&tlng=es (accessed May 27, 2017).

[6] D. Tsolkas, E. Liotou, N. Passas, L. Merakos, A survey on parametric QoE estimation for popular services, J. Netw. Comput. Appl. 77 (2017) 1–17. doi:10.1016/j.jnca.2016.10.016.

[7] M. García-Pineda, J. Segura-García, S. Felici-Castell, A holistic modeling for QoE estimation in live video streaming applications over LTE Advanced technologies with Full and Non Reference approaches, Comput. Commun. 117 (2018) 13–23. doi:https://doi.org/10.1016/j.comcom.2017.12.010.

[8] C.M. Lentisco, L. Bellido, J.C.C. Q, E. Pastor, J.L.A. H, QoE-Based Analysis of DASH Streaming Parameters Over Mobile Broadcast Networks, IEEE Access. 5 (2017) 20684–20694. doi:10.1109/ACCESS.2017.2755438.

[9] S. Malisuwan, D. Milindavanij, W. Kaewphanuekrungsi, Quality of Service (QoS) and Quality of Experience (QoE) of the 4G LTE Perspective, Int. J. Future Comput. Commun. 5 (2016) 158–162.

[10] A. Aloman, A.I. Ispas, P. Ciotirnae, R. Sanchez-Iborra, M.D. Cano, Performance Evaluation of Video Streaming Using MPEG DASH, RTSP, and RTMP in Mobile Networks, in: 2015 8th IFIP Wirel. Mob. Netw. Conf. WMNC, 2015: pp. 144–151. doi:10.1109/WMNC.2015.12.

[11] Z. Yetgin, Z. Göçer, Quality of experience prediction model for progressive downloading over mobile broadcast networks, Telecommun. Syst. 58 (2015) 55–66. doi:10.1007/s11235-014-9873-8.

[12] F. Wamser, D. Staehle, J. Prokopec, A. Maeder, P. Tran-Gia, Utilizing buffered YouTube playtime for QoE-oriented scheduling in OFDMA networks, in: 2012 24th Int. Teletraffic Congr. ITC 24, 2012: pp. 1–8.

[13] A. El Essaili, D. Schroeder, E. Steinbach, D. Staehle, M. Shehada, QoE-Based Traffic and Resource Management for Adaptive HTTP Video Delivery in LTE, IEEE Trans. Circuits Syst. Video Technol. 25 (2015) 988–1001. doi:10.1109/TCSVT.2014.2367355.

[14] A.E. Essaili, D. Schroeder, D. Staehle, M. Shehada, W. Kellerer, E. Steinbach, Quality-of-experience driven adaptive HTTP media delivery, in: 2013 IEEE Int. Conf. Commun. ICC, 2013: pp. 2480–2485. doi:10.1109/ICC.2013.6654905.

[15] M. Solera, M. Toril, I. Palomo, G. Gomez, J. Poncela, A Testbed for Evaluating Video Streaming Services in LTE, Wirel. Pers. Commun. (2017) 1–21. doi:10.1007/s11277-017-4999-0.

[16] M. Torres Vega, D.C. Mocanu, S. Stavrou, A. Liotta, Predictive no-reference assessment of video quality, Signal Process. Image Commun. 52 (2017) 20–32. doi:10.1016/j.image.2016.12.001.

[17] R. Huang, X. Wei, Y. Gao, C. Lv, J. Mao, Q. Bao, Data-driven QoE prediction for IPTV service, Comput. Commun. (2017). doi:10.1016/j.comcom.2017.11.013.

[18] P. Casas, A. D'Alconzo, F. Wamser, M. Seufert, B. Gardlo, A. Schwind, P. Tran-Gia, R. Schatz, Predicting QoE in cellular networks using machine learning and in-smartphone measurements, in: 2017 Ninth Int. Conf. Qual. Multimed. Exp. QoMEX, 2017: pp. 1–6. doi:10.1109/QoMEX.2017.7965687.

[19] S. Satti, C. Schmidmer, M. Obermann, R. Bitto, L. Agarwal, M. Keyhl, P.1203 evaluation of real OTT video services, in: 2017 Ninth Int. Conf. Qual. Multimed. Exp. QoMEX, 2017: pp. 1–3. doi:10.1109/QoMEX.2017.7965682.

[20] I. Gomez, P. Sutton, A. Nag, A. Selim, L. Doyle, V. Ramachandran, A. Kokaram, A software radio LTE network testbed for video quality of experience experimentation, in: 2017 Ninth Int. Conf. Qual. Multimed. Exp. QoMEX, 2017: pp. 1–3. doi:10.1109/QoMEX.2017.7965667.

[21] Amarisoft, (n.d.). https://www.amarisoft.com/ (accessed January 16, 2019).

[22] A. Raake, M.N. Garcia, W. Robitza, P. List, S. Göring, B. Feiten, A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1, in: 2017 Ninth Int. Conf. Qual. Multimed. Exp. QoMEX, 2017: pp. 1–6. doi:10.1109/QoMEX.2017.7965631.

[23] ITU-T Recommendation, P.1203.1 : Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Video quality estimation module, (2017). http://www.itu.int/rec/T-REC-P.1203.1-201710-I (accessed February 2, 2019).

[24] W. Robitza, M.N. Garcia, A. Raake, A modular HTTP adaptive streaming QoE model; Candidate for ITU-T P.1203., in: 2017 Ninth Int. Conf. Qual. Multimed. Exp. QoMEX, 2017: pp. 1–6. doi:10.1109/QoMEX.2017.7965689.

[25] A.D. Clark, Modeling the effects of burst packet loss and recency on subjective voice quality, in: 2001.

[26] W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, K. Yamagishi, M.-N. Garcia, S. Broom, HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P. 1203: Open Databases and Software, in: Proc. 9th ACM Multimed. Syst. Conf., ACM, New York, NY, USA, 2018: pp. 466–471. doi:10.1145/3204949.3208124.

[27] M. Seufert, N. Wehner, P. Casas, Studying the Impact of HAS QoE Factors on the Standardized QoE Model P.1203, in: 2018 IEEE 38th Int. Conf. Distrib. Comput. Syst. ICDCS, 2018: pp. 1636–1641. doi:10.1109/ICDCS.2018.00185.

[28] ITU-T Recommendation, P.911: Subjective audiovisual quality assessment methods for multimedia applications, (1998). https://www.itu.int/rec/T-REC-P.911-199812-I/en (accessed April 12, 2019).

[29] ITU-T Recommendation, P.910: Subjective video quality assessment methods for multimedia applications, (2008). https://www.itu.int/rec/T-REC-P.910-200804-I/en (accessed April 12, 2019).

[30] ITU-T Recommendation, P.1203.2 : Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Audio quality estimation module, (2017). https://www.itu.int/rec/T-REC-P.1203.2-201710-I/en (accessed February 2, 2019).

[31] ITU-T Recommendation, P.1203.3 : Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Quality integration module, (2017). http://www.itu.int/rec/T-REC-P.1203.3-201710-I (accessed February 15, 2019).

[32] D. Ghadiyaram, J. Pan, A.C. Bovik, Learning a Continuous-Time Streaming Video QoE Model, IEEE Trans. Image Process. 27 (2018) 2257–2271. doi:10.1109/TIP.2018.2790347.

[33] D. Ghadiyaram, J. Pan, A.C. Bovik, A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos, IEEE Trans. Circuits Syst. Video Technol. 29 (2019) 183–197. doi:10.1109/TCSVT.2017.2768542.

[34] 3GPP, Specification 36.521-1 : Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Conformance testing, (2016). https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2469 (accessed June 27, 2019).

[35] T. Molloy, Z. Yuan, G.-M. Muntean, Real time emulation of an LTE network using NS-3, in: Limerick, 2014: pp. 251–257. doi:10.1049/cp.2014.0694.

[36] H.-F. Bermudez, R. Sanchez-Iborra, J.L. Arciniegas, W.Y. Campo, M.-D. Cano, Statistical validation of an LTE emulation tool using live video streaming over reliable transport protocols, Telecommun. Syst. 71 (2019) 491–504. doi:10.1007/s11235-018-0521-6.

[37] H.F. Bermudez, R. Sanchez-Iborra, J.L. Arciniegas, W.Y. Campo, M.D. Cano, Performance validation of NS3-LTE emulation for live video streaming under QoS parameters, in: 2017: pp. 300–307. doi:10.1109/WiMOB.2017.8115836.

[38] Wowza, (2019). https://www.wowza.com/products/streaming-engine (accessed February 8, 2019).

[39] Wireshark, (2018). https://www.wireshark.org/ (accessed March 5, 2019).

[40] M. Nakagami, The m-Distribution, a general formula of intensity of rapid fading, Stat. Methods Radio Wave Propag. Proc. Symp. Held Univ. Calif. (1960) 3–36.

[41] NS-3, Propagation Models, (2019). https://www.nsnam.org/docs/release/3.18/doxygen/group__propagation.html (accessed March 2, 2019).

[42] ITU-T Recommendation, P.913 : Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, (2016). https://www.itu.int/rec/T-REC-P.913 (accessed April 15, 2019).