QoE estimation for different adaptive streaming techniques in mobile networks

Estimación de la QoE para diferentes técnicas de streaming adaptativo en redes móviles

H. F. Bermúdez-Orozco; W. Y. Campo-Muñoz; E. Astaiza-Hoyos

DOI: https://doi.org/10.22517/23447214.24746

Abstract—Video services are becoming more and more popular for mobile network users and require greater and greater resources and provisions from telecommunications service providers. But operators suffer from problems of interoperability between the different adaptive transmissions techniques they employ in an attempt to satisfy the quality of experience (QoE) of the service provided to users and improve network performance. This article presents a comparison of four such streaming techniques - DASH (dynamic adaptive streaming over HTTP), HDS (HTTP dynamic streaming), HLS (HTTP2 live streaming) and HSS (HTTP smooth streaming) - used in a live video playback by a user in different test scenarios on an emulated long-term evolution (LTE) network. Comparison of performance was carried out using the mean opinion score (MOS) metric calculated based on ITU-T Recommendation P.1203. This recommendation focuses on bitstream-based parametric quality assessment of progressive downloads and adaptive audiovisual streaming services through a reliable transport protocol. From the results obtained in the evaluation of the different adaptive transmission techniques for the test scenarios presented, it is observed that the DASH protocol presented a better performance in the QoE evaluation using the MOS metric compared to the other protocols analyzed.

Index Terms—Adaptive Transmission, LTE, MOS, QoE, Video.

Recent years have seen an increase in the number of users using video streaming services to play content that is either live (LVS - live video streaming) or on demand (VoD - video on demand). The number of users and mobile devices (e.g., laptop, smartphone, tablet) that access these types of service through wireless mobile networks is also increasing [1]. Moreover, the wide variety of devices offered by the market increases the variability of characteristics between devices: screen size or resolution, type of internet connection, contracted bandwidth, and network status at the time of video playback, etc.

Success in video streaming focuses on the user being able to display content on their device with the minimum of failure or delay. The service provider is thus obliged to undertake certain network management tasks - the monitoring and control of bandwidth, delay, jitter, throughput and packet loss - that ensure an adequate level of quality for users. These tasks are much more complex in a wireless environment where difficulties include wireless signal coverage, a high rate of packet loss, and instability of the wireless channel, these being the result of phenomena pertaining to the channel itself, such as multipaths, fading, interference, and noise [2], [3], [4]. These can considerably affect the performance of wireless mobile networks and bring down the quality of experience evaluation [5].

Video services have undergone a great transformation in recent times with respect to protocols and techniques used for their transport. In early versions of video delivery services, UDP (user datagram protocol) was used as transport protocol due to its simplicity and the reduced amount of control traffic.
Later, with the improvement of data networks, the reliable TCP (transmission control protocol) was adopted and, today, the most popular video services on the Internet, for example, VoD or LVS use adaptive transport techniques in preference to Hypertext Transfer Protocol (HTTP) [6]. These techniques adapt the speed of video data transmission to the network available bandwidth, so that even while in progress they adopt the most suitable video encoding settings given the conditions of end user and data path [7]. The following may be considered among the most popular video adaptation techniques: (i) HTTP smooth streaming (HSS) from Microsoft [8], (ii) HTTP2 live streaming (HLS) from Apple [9], (iii) HTTP dynamic streaming (HDS) from Adobe [10], and finally (iv) dynamic adaptive streaming over HTTP (DASH), from the Moving Picture Expert Group (MPEG), defined as an ISO/IEC standard (23009 -1: 2012) [11].

Regarding the use of video streaming techniques in wireless mobile networks, LTE technology since its inception has progressively incorporated improvements in response to the increase in mobile network traffic, seeking to enhance the video experience for its end users. These enhancements support features such as fast initial startup, bandwidth efficiency, adaptive bit rate switching, adaptation to content delivery network (CDN) properties, reuse of HTTP servers and caches, reuse of existing media playback engines, support for delivery of services on demand, live and displaced in time and simplicity for wide adoption [12].

Meanwhile, for video quality measurements, two contexts can be differentiated - subjective quality measures and objective quality measures [13]. As regards the former, the International Telecommunications Union (ITU-T) has formalized a number of methodologies for their assessment in several recommendations, among which are ITU-T P.910 and P.911, [14], [15], which look to obtain the average quality rating of all users for a particular video transmission. This rating is known as the mean opinion score (MOS). Objective quality metrics, on the other hand, are algorithms designed to characterize video quality and predict the MOS of users. These metrics are not based on service user surveys but on parameters measured within the network or video stream itself. Thus, Recommendations P.1201 (ITU-T, 2012), P.1202 [16] and P.1203 [17] provide an overview of models for non-intrusive quality monitoring based on the IP protocol by analyzing packet header information and video signal flow. In this research, Recommendation P.1203 was used since it focuses on evaluation of parametric quality based on video streams of progressive downloads and adaptive audiovisual transmission services through a reliable transport protocol such as TCP. The P.1203 standard meanwhile is one of the first standardized QoE estimation methodologies to incorporate machine learning techniques for QoE prediction. The aim of this study therefore is to estimate QoE of the main adaptive streaming techniques for a particular service, in this case LVS, over 4G networks emulated using Network Simulator-NS3 software. The main contributions of the work are as follows: (i) Compare performance of the most outstanding current adaptive streaming techniques using the P.2103 standard and (ii) analyze which technique performs best for the LVS service in an emulated 4G network.

The paper is organized as follows. In Section II, studies related to techniques of sending video through adaptive streaming are analyzed. Section III presents the methodology. In section IV, the results and their discussion are found and finally, Section V presents the conclusions and identifies future work.

II. RELATED WORK

The following presents an analysis of the work reviewed on this topic. In [18] the authors studied the effect of various QoS parameters on the estimation of QoE, for which they presented the assessment of three video streaming protocols, MPEG-DASH, RTSP and RTMP (Real Time Messaging Protocol), for VoD and LVS services over 4G and WiFi technologies. In the experiments carried out for the assessment, the Rohde & Schwarz CMW500 broadband communications analyzer and smartphones were used. Important contributions of the work included the study of comparative performance of adaptive and non-adaptive video streaming protocols under real network scenarios in terms of QoE and the comparative assessment of two parametric models to evaluate QoE according to the ITU-T R.1201 recommendation. All results were obtained from measurement campaigns in a real test scenario, making it impossible to vary network parameters that can be decisive in QoE evaluation. The authors of [19] analyzed the performance of different adaptive video transmission transport protocols, namely DASH, HLS, HSS and HDS, in different test scenarios with real situations of video signal cuts using the Net.Storm impairment generator [20]. The comparative analysis focused on studying the behavior of different players, both from TV providers and standard players, obtaining conclusions about the robustness of each of the adaptive streaming techniques analyzed. The disadvantage of this study was that the authors for the construction of the experimentation scenarios used a CDN service provider company, making it difficult to manipulate the initial configuration parameters of the services used. In [21], an assessment of network architectures used by commercial companies was carried out, adaptive streaming techniques (DASH, HLS, HSS and HDS) were studied and a case study was designed for a network of video streaming for telemedicine applications between the hospitals of three small coastal cities and the Hospital de la Seguridad Social in Guayaquil, Ecuador.

The authors of [22] presented a CDN test bed using HTTP adaptive transmission technologies. The assessment results were calculated based on QoS parameters (bandwidth, delay and packet loss), obtaining the performance of the use of adaptive streaming techniques with CDN. As the main conclusion, the authors showed that the use of these types of adaptive technologies improves the performance of server load and reduces network congestion, thus providing a better experience for end users. However, the authors did not specify the type of streaming technology used. In [23], the authors offered an overview of the state of the art of adaptive streaming
techniques over HTTP through multimedia domains and different networks. Results obtained were shown to analyze the challenges and solutions in adaptive transmission algorithms, QoE improvement, network protocols, buffering, etc. The authors also focused on various challenges about the factors that influence QoE in a variable network condition. The authors of [24] proposed a virtual test bed for the implementation of adaptive video streaming experiments. The authors studied different parameters that affect QoE performance, among which are initial delay (start delay at the beginning of a video playback), changes in video encoding, frequency switches (the number of times quality is changed), accumulated video time (the number and duration of video stop events - stallings), CPU usage, and battery power consumption. Furthermore, the authors found a relationship of these parameters with a subjective estimate of QoE. To estimate QoE, original and received videos on the virtual benchmark were evaluated by expert users. As a contribution to the body of work reviewed in this section, this research presents the comparison of the different adaptive streaming techniques most widely used among end users, through QoE estimates, using for the first time the methodology proposed in Recommendation P.1203 in a 4G (LTE) wireless environment.

III. METHODOLOGY

An emulation scenario was built in order to estimate QoE based on ITU-T Recommendation P.1203 and carry out a performance comparison of the DASH, HDS, HLS and HSS adaptive streaming techniques under real operating conditions. The scenario was previously proposed in [25] and then adapted and validated in [26] and [27]. This scenario is made up of three personal computers (PC), the characteristics of which can be seen in Table I. Fig. 1 shows the diagram of deployment of the emulation system.

<table>
<thead>
<tr>
<th>PC</th>
<th>Characteristics</th>
<th>Operative system</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC -1 LTE emulated in NS3</strong></td>
<td>Intel Core 2 at 2.13 GHz, with 4 GB of RAM</td>
<td>Linux Ubuntu 14.04 LTS</td>
<td>NS3.26</td>
</tr>
<tr>
<td><strong>PC -2 Server</strong></td>
<td>Intel Core i7-3612QM CPU 2.1GHz, 8 GB of RAM</td>
<td>Windows 7 Professional</td>
<td>Wowza streaming engine, Wireshark</td>
</tr>
<tr>
<td><strong>PC -3 Customer</strong></td>
<td>Intel Core 2 Duo CPU 2.1GHzx2, 4 GB of RAM</td>
<td>Windows 7 Professional or Linux Ubuntu</td>
<td>VLC, Akamai, Adaptive media player, Wireshark</td>
</tr>
</tbody>
</table>

PC1 hosts the 4G-LTE network emulated with the LENA library on NS3. This LTE network is composed of a remote host, a SGW/PGW node (Serving Gateway/Packet Data Network Gateway), an eNB base station (evolved NodeB) and an end user equipment (UE) that acquires the role of mobile device. Real video traffic thus reaches the UE node, injected into the system by the LVS Wowza streaming server (PC2) through the emulated LTE network. Since within the emulated system the UE consumes the video traffic but does not view it, PC3 is used for the reproduction and monitoring of the information received by the virtualized UE node. The HIL (Hardware in the Loop) platform is used to enable communication between PC2-Host-Remote and UE-PC3, allowing PC1 to be taken as a black box with virtualized components that receives and delivers data to real end systems, PC2 and PC3.

Three test scenarios were defined in order to carry out performance comparison of the different adaptive streaming techniques: (1) a user located 30m from the eNB, for whom the different network delay times {0, 25, 50, 75, 100, 125 and 150 ms} were varied in a controlled way, as were different

![Deployment diagram of the emulation system](image-url)
percentages of packet loss \( \{0, 0.5, 1, 2, \text{ and } 3\%\} \). This scenario is considered as a static scenario; (2) a UE that moved away from the eNB with uniform speed and direction, for which different speeds \( \{1, 1.5, 2, 2.5 \text{ and } 3 \text{ m/s}\} \) are varied in a controlled way, while direction follows a straight line \( y = x \); and finally, (3) a UE that moved away from the eNB with random speed and direction. Scenarios (2) and (3) were considered as dynamic scenarios.

For each of the test scenarios, ten tests of 180s each of the video clip "Big Buck Bunny" were transmitted live, consumed by the customer using the emulated 4G network (LTE). At the customer, the parameters necessary to estimate QoE were extracted using the methodology proposed by the P.1203 standard. The parameters were encoding settings (see Table II) and initial buffer and stalling events (quantity, start time, duration). A script developed in Matlab® was used to estimate QoE.

**Table II**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coding adjustments</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Video coding</strong></td>
<td></td>
</tr>
<tr>
<td>Coder</td>
<td>H.264</td>
</tr>
<tr>
<td>Resolutions (px)</td>
<td>240p (426x240); 360p (640x360); 480p (854x480); 720p (1280x720); 1080p (1920x1080)</td>
</tr>
<tr>
<td>Bit coding rate (Kbps)</td>
<td>528 for 240p; 878 for 360p; 1128 for 480p; 2628 for 720p; 4628 for 1080p</td>
</tr>
<tr>
<td>Frames per second</td>
<td>30</td>
</tr>
<tr>
<td><strong>Audio coding</strong></td>
<td></td>
</tr>
<tr>
<td>Channels</td>
<td>Stereo</td>
</tr>
<tr>
<td>Bitrate (Kbps)</td>
<td>96 for 240p; 128 for 360p, 480p; 192 for 720p, 1080p</td>
</tr>
<tr>
<td>Sample frequency</td>
<td>44.100 kHz</td>
</tr>
</tbody>
</table>

**IV. RESULTS AND DISCUSSION**

Below, a comparison is found of the four HTTP adaptive streaming techniques for the different test scenarios. Comparisons were made based on the results obtained for the QoE estimation from the MOS, the assessment scale of which was the following: bad \(1-1.9\), poor \(2-2.9\), fair \(3-3.9\), good \(4-4.9\), and excellent \(5\) \[15\]. The results presented correspond to the mean value of ten tests carried out for each of the test scenarios. QoE is calculated by applying the methodology in the P.1203 standard. All measurements were made for the provision of the LVS service in the emulated 4G network described above.

Figs. 2-5 present the QoE estimate from MOS calculations for the different test scenarios. Fig. 2 shows for Test Scenario 1 how DASH and HLS adaptive streaming techniques performed best when controlled delays of less than or equal to 50 ms are introduced to the emulated system, with HSS and HDS doing less well. It can be seen that, after 75 ms of delay, the QoE estimate for the different techniques studied is considered poor, since QoE is located in the qualification range 2.9. Delay values of up to 50 ms for the DASH technique can be considered as acceptable on obtaining a QoE of at least fair (MOS = 3) for the LVS service in the 4G system studied.

In Fig. 3, for test scenario 1, from a QoE perspective it can be seen that the performance of the DASH adaptive streaming technique was superior to the performance obtained with the other streaming techniques when different packet loss values were introduced in a controlled way. For the DASH technique, it is clearly seen that for values less than 0.5% packet loss, the QoE estimate through MOS is greater than 4, considered a good evaluation, i.e. in the assessment range: 4-4.9. On the contrary, for values greater than 0.5% packet loss, QoE evaluation was considered as bad, i.e. in the assessment range: 2-2.9, rendering the service unviable from the user point of view.

Fig. 4 presents the results obtained for test scenario 2. It can be seen that the technology with the best performance in terms of user mobility was DASH. With this technique, estimation of QoE was good, since it is in the rating range of 4-4.9 for speeds between 1-2.5 m/s. Only for the speed of 3m/s was the evaluation of QoE estimate fair, i.e. in the evaluation range 3-3.9. For the other technologies, HLS, HDS and HSS, it can be seen that the performance measured by estimating QoE was fair (evaluation range: 3-3.9) for speeds of 1 to 2.5 m/s. Only HSS
technology for the speed of 3m/s presented a poor QoE estimate, i.e. in the range: 2-2.9.

Fig. 4. Estimated QoE in scenario 2 (UE moving away from eNB with uniform speed and direction).

Fig. 5 shows the results obtained for test scenario 3. It was observed how the DASH technique again returned the best performance (MOS on the scale 4-4.9) in terms of QoE when the user moves at a random speed between 1 - 5m/s and in a random direction within an 80m x 80m rectangle. The rest of the technologies obtained a fair QoE estimate (evaluation range 3-3.9).

Fig. 5. Estimated QoE in scenario 3 (UE moving away from the eNB with random speed and directions).

Figs.6-10 show results obtained for stalling events for different test scenarios, as defined in the P.1203 standard as having a high impact on QoE estimation [24]; Fig. 6 and 7 show duration and quantity of stalling events for test scenario 1. It was observed, in Fig. 6, how as controlled delays greater than 100 ms are introduced in the emulation system, a longer duration of stalling events is encountered in DASH and HDS techniques. On the other hand, the number of stalling events (see Fig. 7) increases in a less abrupt way as delay time increases for the HDS and HSS technologies. In the case of DASH, the number of stalling events reaches its maximum value (approx. 4) when the delay was of the order of 100 ms, and then decreases slightly. However, this does not mean that DASH performance improves from 100 ms to 125 or 150 ms. On observing the duration of the stalling events for DASH, it can be seen that, although there were fewer stalling events, their duration was much longer, even exceeding 30 s. This means that of the 180 s of video playback, 99 s is stopped. As a result, QoE estimate is very low for DASH in delays above 100 ms, as evidenced in Fig. 2, where a QoE estimate of bad was observed (evaluation range: 1-2). In conclusion, it can be stated that for DASH and HDS streaming techniques, the higher the delay, the fewer stalling events but the longer the duration. While for HLS and HSS, the more delay the more stalling events but shorter.

For test scenario 2, where the UE moves away from the eNB with uniform speed and a single direction, it was observed in Figures 8 and 9 that the duration and number of stalling events affect HDS and HSS technologies much more than with DASH and HLS. In addition, DASH presents fewer stalling events than HLS and has a shorter average duration, see Fig. 9. Comparing with the QoE measurements, DASH and HLS obtained better results with a MOS of above 3.25 for all speeds.

Fig. 6. Duration of stalling events for emulation scenario 1.

Fig. 7. Number of stalling events for emulation scenario 1.
Fig. 8. Duration of stalling events for emulation scenario 2.

Fig. 9. Number of stalling events for emulation scenario 2.

Fig. 10 shows the results obtained for test scenario 3. A behavior very similar to scenario 2 was observed, where the DASH technique presented a better performance in terms of quantity and duration of stalling events.

Meanwhile, with respect to the initial buffer parameter, Fig. 11 shows its duration for scenario 2, when different values of percent packet loss were entered in a controlled manner. It was observed that, as the percentage of packet loss increases, the duration of the initial buffer increases, and is greater for the DASH and HLS technologies than for HDS and HSS. By associating these results with the QoE estimates, it could be inferred that for scenario 2, the use of a larger initial buffer is more favorable in terms of QoE. Similar behavior of the initial buffer parameter was observed in test scenarios 1 and 3; as well as when the delay or speed was varied in a controlled fashion.

V. CONCLUSIONS

A comparative study of the performance of different video streaming techniques based on QoE was carried out through a controlled variation of such QoS parameters as delay and packet loss in different static and dynamic experimentation scenarios defined within the research. Regarding the delay parameter, the following observations were made: (i) the value of 50 ms was considered to be the maximum limit to ensure quality of LVS service in 4G (LTE) networks in dynamic and static scenarios. Furthermore, 50 ms was considered quite a demanding value, since the reference [29] values given for delay range from 0 to 400 ms. (ii) The DASH adaptive streaming technique outperformed the other techniques studied, a behavior that was most evident in the dynamic scenarios. (iii) Regarding the stalling and initial buffer events for the static and dynamic scenarios, it was evident that as delay increased, the duration and number of stalling events increased for all the adaptive streaming techniques studied. The strongest performance with this parameter was again observed for the DASH technique. Finally, on controlling variations in the packet loss parameter for the defined static and dynamic scenarios, the following observations were made: (i) for all the scenarios and tests carried out under the specific operating parameters, the adaptive streaming that performed best with respect to QoE (MOS) was once again that of DASH. (ii) It was observed that on exceeding 0.5% packet loss, the LVS service was rendered unviable, from the point of view of the user, given that for these values of percent packet loss, the QoE assessment achieved an evaluation only of bad (2-2.9) or lower. According to the results obtained,
the use of the DASH protocol may become the solution to the different interoperability problems that telecommunications operators present due to the use of different adaptive streaming techniques. Currently, the authors are conducting research to establish relationships between different QoS metrics and the objective estimate of QoE for the streaming service; additionally, this research will allow extending studies and emulation scenarios to other access technologies, such as WiFi networks, LTE-A Pro and 5G.

ACKNOWLEDGMENT

Very special thanks are due to the Telematics Engineering Research Group - GIT of the Universidad Politécnica de Cartagena, in Spain, where the emulation scenario that provided the starting point for carrying out this research was implemented, and in particular thanks to Dr. María Dolores Cano B. and Dr. Ramón Sánchez Ibora. In addition, we are very grateful to Mr. Colin McLachlan for suggestions related to the starting point for carrying out this research was Cartagena, in Spain, where the emulation scenario that provided the objective estimate of QoE for the streaming service; establishing relationships between different QoS metrics and the objective estimate of QoE for the streaming service; additionally, this research will allow extending studies and emulation scenarios to other access technologies, such as WiFi networks, LTE-A Pro and 5G.

REFERENCES

H. F. Bermúdez-Orozco, Author, is a Titular Professor and Researcher at the Electronic Engineering Program in the University of Quindío, Colombia. From the University of Cauca (Colombia), he received the degrees in Electronics and Telecommunication Engineer in 2000, Masters in Electronics and Telecommunications in 2010 and PhD in Telematic Engineering in 2020, where he received a laureate mention for his doctoral research thesis. He made a doctoral research stay at the Polytechnic University of Cartagena UPCT in Cartagena Murcia (Spain) in 2018. He researcher of the Telecommunications Research group (GITUQ) at Universidad of Quindío. Areas of interest: wireless communications, radiant systems and propagation, modeling of traffic of telematic services, Quality of Service (QoS)/Quality of user Experience (QoE).
ORCID: https://orcid.org/0000-0002-8101-3764

Orcid: https://orcid.org/0000-0003-2706-0962

W. Y. Campo, Author, PhD in Telematic Engineering, Magister in Engineering, Telematic Area and Engineer in Electronics and Telecommunications from the University of Cauca Colombia. He is a researcher at the GITUQ group of the University of Quindío Colombia, where he is currently linked as a professor. His fields of interest are: Software Defined Networking SDN, 5G wireless networks and teletraffic.
Orcid: https://orcid.org/0000-0001-8585-706X