Mind the (QoE) Gap: On the Incompatibility of Web and Video QoE Models in the Wild

Michael Seufert*, Nikolas Wehner*, Viktoria Wieser*, Pedro Casas†, Germán Capdehourat‡

*University of Würzburg, Würzburg, Germany
†AIT Austrian Institute of Technology, Vienna, Austria
‡Plan Ceibal, Montevideo, Uruguay and Universidad de la República, Montevideo, Uruguay
{michael.seufert|nikolas.wehner}@uni-wuerzburg.de, viktoria.wieser@informatik.uni-wuerzburg.de
pedro.casas@ait.ac.at, gcapdehourat@ceibal.edu.uy

Abstract—Education Service Providers (ESPs) have a paramount role in the digitization of education, providing reliable devices for students and teachers and high quality Internet access at schools. In this paper, a large-scale, passive, in-device Quality of Experience (QoE) monitoring system is presented, which was deployed into a nationwide network of education-purpose devices. Four months' worth of continuous measurements were conducted by an ESP, covering more than 800 education centers and about 4000 devices, used both in schools and at home. When analyzing the QoE of web sessions in school networks, we identify a fundamental issue with the compatibility of web browsing and video QoE models, which inhibits the successful application of QoE-aware network management for multiple services.

Index Terms—QoE Models; Distributed Network Monitoring; Web Browsing; Video Streaming; Educational Networks; Network Measurements; QoE-aware Network Management

I. Introduction

The omnipresence and paramount role of the Internet in our daily lives have paved the way to a new paradigm in formal school education. Nowadays, education-specific digital contents are remotely accessed by the students, and teachers rely on specific online platforms to transform the overall learning experience. In this new Internet-supported educational paradigm, the role of the Education Service Providers (ESPs) is indispensable. ESPs are usually for-profit or non-profit organizations which work with the national education systems to help them implement comprehensive reforms towards digitization. Although the role of ESPs is not the same in all countries and regions, the different programs typically include at least two key components: providing devices for students and teachers (either laptops or tablets) and deploying Wi-Fi infrastructure at schools for Internet access (through a local ISP connection). Thus, it is of major importance for the ESPs to guarantee high quality services, as performance degradation would impact the quality of education.

In analogy to Internet Service Providers (ISPs), who rely on Quality of Experience (QoE) monitoring to analyze their networks performance, ESPs need similar systems to assure a high QoE to support teachers and students during the learning experience. In this respect, ESPs have the same lack-of-visibility problems as ISPs, facing the challenge to properly measure QoE from pure network traffic monitoring

and analysis. A main advantage of the ESPs is that, in many cases, they also provide the end user devices. Thus, they can easily deploy application-layer QoE monitoring systems within these devices, providing an augmented degree of visibility into the activity, behavior, and performance of different applications. Depending on the users location, there are two different situations that are relevant for ESPs. At schools, the QoE would depend on two factors managed by the ESPs, the device and the Wi-Fi Internet access. On the other hand, when the users are at home, they may still use the devices provided by the ESP, but in that case they rely on their own Internet connection. Nowadays, the latter is usually also a Wi-Fi access, but the backbone can be a fixed or mobile connection, from the same or other ISPs as in schools.

In collaboration with Plan Ceibal [1], a major ESP which leads a nationwide one-to-one computing program in Uruguay, we implemented and deployed a passive web QoE monitoring system directly within the end-devices. Through this monitoring system, we collected four months of continuous, nationwide measurements covering more than 800 education centers and about 4000 devices from teachers and students, who may use them both in schools and at home, and both for education and entertainment. Now in 2020, the situation has changed dramatically due to COVID-19 pandemic, with an increased relevance of home-schooling, where users rely on their own home Internet connections. However, the study was conducted prior to the COVID-19 outbreak, so the measurements were conducted in a time of regular face-to-face school classes.

To the best of our knowledge, this is the first study focusing on passive QoE monitoring in ESP networks. The key performance indicators (KPIs) of web browsing and video QoE were monitored, which allow to quantify the QoE of the visited webpages and the played out videos. This rich data set helped us to identify a mismatch between web and video QoE models when trying to mix them both into a single session QoE score, something that inhibits the successful application of QoE-aware network management for multiple services.

Therefore, this work is structured as follows. Section II describes related works on web and video QoE. Section III outlines the implemented monitoring system and describes the dataset. Section IV presents the QoE results for web sessions and details the incompatibility between web and video QoE

models, before Section V concludes the work.

II. RELATED WORK

According to a widely accepted definition [2], Quality of Experience (QoE) of a multimedia system is influenced by context, user, system, and content level factors. With respect to system factors of web browsing, response times were identified as the most important QoE factor [3]. Thus, the first web QoE models were based on page load time (PLT), e.g., [4]. A refined approach is the "above-the-fold" (ATF) time, i.e., the time until the visible portion of a web page has been fully loaded, which could also be included in traditional web QoE models [5]. Additionally, integral-based metrics were proposed, such as Google's SpeedIndex, which take into account the whole visual progress of the page loading. As SpeedIndex takes into account the rendered pixels, it can only be measured on application-level. Thus, [6], [7] developed approximations for the SpeedIndex, which can be measured in the network.

For video QoE, most works on (adaptive) video streaming agree that initial delay, stalling, and quality adaptation are the most dominant QoE factors [8]. Stalling, i.e., playback interruptions due to buffer depletion, is considered the worst QoE degradation [9], [10], and should be avoided. Furthermore, video streams should be played out with high visual quality [11]. In contrast, initial delay has only a small impact on the QoE [4]. A huge variety of QoE models was proposed in literature, e.g., [8], which cumulated in the recent standardization of the P.1203 model [12], [13], which will also be used in the context of this work.

III. METHODOLOGY

The passive QoE monitoring system is based on a Chrome browser plugin, as this is the most popular browser worldwide - and in particular in Uruguay, with a share of more than 80% according to [14]. The plugin is composed of two JavaScript files, content and background, and a manifest file in JSON format (for plugin permissions and links to both scripts). The content script is injected into every webpage and retrieves the browsing timestamp, the URL, and page load timing information from the PerformanceNavigationTiming API [15].

If a video element is embedded into the webpage, the content script will periodically log the video streaming progress every 250ms. Therefore, the video element is queried and the relevant streaming state (e.g., current playtime, buffer, player state, video resolution, video id, screen resolution, advertisement clips) is extracted, and written to a verbose, YoMoApp-style log [16]–[18], which allows to compute video QoE metrics (e.g., initial delay, stalling events, resolution changes) in a postprocessing step. Finally, the logged data are sent to the background script either if a new webpage has to be monitored, or if the webpage is closed.

The background script is running in the browser during the whole browsing session, where it creates and stores an anonymized user ID. Additionally, it listens if a new webpage was browsed via pushState-based navigation [19] and sends a message to the content script to start QoE monitoring. Note that in case of pushState-based navigation, the performance element is not updated as the page is not loaded in the classical sense, but only its content is altered. Nevertheless, in this case, the timestamp of the content update and the new URL are recorded. Finally, the background script receives the monitored information from the content script and stores it locally to a log in the browser storage, which is then sent to a centralized server. This centralized web server collects all the monitored QoE information, storing the web browsing logs as well as additional meta information, such as timestamps, user ID, IP address, browsed URL, etc. Afterwards, browsing metrics are merged with the meta information and inserted into a database.

Moreover, the background script analyzes the verbose, YoMoApp-style video logs in order to compact the streaming information and to extract video QoE metrics. This avoids the need to store the full video logs, which considerably reduces the storage consumption. Afterwards, browsing and video metrics are each merged with the meta information and inserted into separate tables. A small share (0.1%) of full video logs is nevertheless stored in another table to also keep the detailed insights into the video streaming and the resulting QoE for some of the streamed videos.

The QoE monitoring system was deployed in a real-world educational scenario, installing the described plugin in the laptops provided by Plan Ceibal [1]. We shall recall that devices are handed over to teachers and students, and they are used both in schools and at homes. Concerning the ethical dimension of this study, it is worth to mention that it was approved by the Plan Ceibal ethical and data privacy committee. The users gave their consent to collect the anonymized data from their devices, which was handled according to the Uruguayan and European privacy protection legislation.

The QoE measurement campaign was carried out during the last four months of 2019, which corresponds to the end of the school year in Uruguay, and also the time of greatest network usage at schools [20], [21]. The massive plugin deployment was done through the devices software update system managed by Plan Ceibal. The data collection process ended up in a dataset with 5,641,034 records corresponding to 3,887 unique devices. As updates of the PerformanceNavigationTiming API are not always properly recorded by the browser when a new page is browsed, a lot of pages report page load information of previous pages. This may happen in particular when a subpage with the same domain as the previous page is accessed. These duplicate page load data had to be filtered, which resulted in 2,654,634 QoE measurements for web pages. Another dataset for video content was collected, only for the cases when users actually played out a video in their browsers. This dataset has 678,549 records from 3,258 unique devices, and as expected, most of the users with web navigation data also have video data (90%), given the current popularity of such content.

The data was collected from 84,772 different IP addresses, of which only 818 (1%) correspond to schools, but they account for almost 15% of the records in the dataset. All schools have high-end Wi-Fi Internet access and the backbone
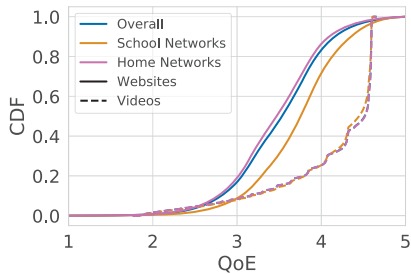
Fig. 1: Web and Video QoE distributions.

Fig. 2: Session QoE distributions.

is an ISP broadband optical fiber connection. The dataset collected for video playback QoE includes information from 65,531 different IP addresses, of which only 685 (1%) are from schools. The total number of video QoE records corresponding to Plan Ceibal sites is 153,935 (almost 7% of the total). We shall recall that in all cases, for both web browsing and video, the devices that collected the data were the laptops provided by Plan Ceibal to students and teachers, so all the analysis carried out in our study corresponds to the same type of device.

IV. QoE RESULTS FOR WEB SESSIONS

To investigate the QoE of web sessions, two standard and/or state-of-the-art QoE models for web browsing and video streaming are utilized. To evaluate the QoE of a single page load, the WQL PLT model from [5] is used: $QoE = -0.5368 \cdot \log(PLT) + 7.9035$. For videos, QoE scores were computed using the standardized P.1203 model [12], [13]. As only metadata could be monitored, mode 0 of P.1203 was used, which considers bitrate, frame rate, and resolution. Both models map to a QoE score on a continuous scale, where 1 indicates bad QoE and 5 indicates excellent QoE.

Figure 1 shows the CDF of the QoE for all monitored websites in all networks (blue), as well as for websites accessed from school networks (orange) and home networks (violet) separately. For this, the QoE of websites is depicted with a solid line, and the QoE of videos is depicted with a dashed line. Overall, it can be seen that only 3.7% of the websites had a score below 2.5, which indicates poor or worse QoE. On the other hand, 49.4% of the websites could be browsed with good or better QoE, i.e., a QoE score above 3.5. Nevertheless, only 2.9% of the pages had an excellent QoE score above 4.5. As a consequence, 44% of the websites showed a QoE between a 2.5 and 3.5. Since the orange solid CDF is right of the violet solid CDF, a clear tendency can be observed that browsing in school networks (orange) gives a better web QoE than browsing in home networks (violet). A possible reason to explain this difference is that the Wi-Fi access at schools corresponds to a planned network with high-end equipment, while at homes the Wi-Fi is usually of poorer quality and without planning (e.g., consumer-grade WiFi routers with no channel and Tx power management to reduce interference between neighboring APs).

When it comes to video streaming, it can be seen that there is almost no difference with respect to the network, since all three dashed CDFs are very close to each other and even
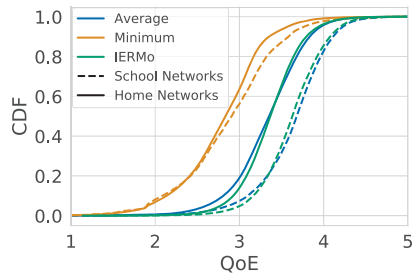
overlap. Overall, the plot shows that only 4.6% of the videos had a poor or worse QoE (< 2.5), while 33.6% of the videos had good or better QoE (> 3.5). 51.1% of the videos could even be streamed with an excellent QoE score (> 4.5). Here, a clear discrepancy between the QoE models for web browsing and video streaming can be observed, as it is very unlikely that, e.g., for the same school networks, videos reach a much higher QoE compared to websites. This interesting finding will be investigated in more detail below.

Next, the session QoE is investigated. For this, the single page visits have to be mapped to sessions first. As proposed in literature [22]–[24], a 30 minutes threshold as think-time is used. This means that a web session ends if a user does not request a new web page within 30 minutes after the last web page request. This classification approach results in 209,020 different sessions by 3,887 users in total, i.e., every user initiated around 53 sessions on average. For each session, two simple session QoE scores are computed as the average and the minimum of the QoE scores of all websites and all videos within that session. However, a critical aspect in session QoE modeling is the memory effect, which occurs when previously experienced stimuli impact the perception of the current stimuli. To account for the memory effect, the Iterative Exponential Regression Model (IERMo) proposed in [25] is also applied to the QoE scores within a session.

The original IERMo model was developed for page load times only, thus, it was slightly adapted to be able to use QoE scores directly. The adapted model takes a sequence of QoE scores as input and estimates the mean opinion score (MOS) for different user groups. In particular, the model considers the short-term memory effect only, i.e., the computation of the current MOS depends only on the previous MOS. The function of the model is not time-dependent, but uses an exponential regression to model a shrinking decay over consecutive stimuli with similar intensity. When a stimulus with different intensity occurs, the decay is reset to a high weight for the MOS computation of the current stimulus. The required similarity is computed by comparing the differences of the QoE scores with a threshold $\epsilon = 0.3$, which is an appropriate threshold for noticeable differences. The final equation for the adaptation of the QoE scores is shown in Equation 1, where w corresponds to the sensitivity of the user and j indicates how often a user perceived a similar stimulus in a row, with $j = 1$ if a new stimulus has been perceived.

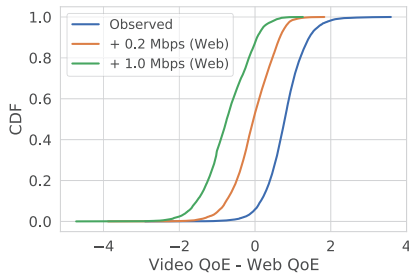$$MOS_i = MOS_i - we^{-j} \cdot (MOS_{i-1} - MOS_i) \quad (1)$$

Fig. 3: Discrepancy between QoE models.

By using different values for w, the sensitivity of the user can be modeled, where $w = 0$ represents no sensitivity, i.e., the user does not care about the QoE score fluctuations, and higher w represent higher sensitivity to fluctuations. As proposed in the original paper [25], $w = 0.254$ will be used here. The IERMo-updated QoE scores are then averaged to the overall IERMo session QoE score.

The CDFs of the average session QoE scores (blue), the minimum session QoE scores (orange), and the IERMo session QoE scores (green) are depicted in Figure 2. Note that when analyzing the temporal characteristics of the session QoE scores, no strong variations over the course of the day could be found in the observed data, such as effects of peak usage hours. Thus, only CDFs are investigated here. It can be seen that, in the context of school networks, the average QoE (blue dashed) does not drop very low, since only 0.7% of the sessions have a QoE score below 2.5. This means that there are very few sessions which have a constantly low QoE for all of their websites and videos. Also to the other extreme, there are only 0.8% of the sessions, which have a QoE score above 4.5, i.e., a constantly excellent QoE, which shows that there is still some room for improvements by network management. Both effects also benefit from using the average QoE as a simple session QoE model, which will smooth extreme QoE scores. In contrast, in home networks, the average QoE (blue solid) is significantly shifted to the left, i.e., to lower QoE. As observed in Figure 1, this effect is due to the lower QoE of webpages.

The CDFs of minimum session QoE (orange) are located more to the left, showing a similar shape until the median at around 2.9 for both school and home networks. Then, the CDF of school networks shows slightly higher minimum QoE scores, which is due to the generally higher QoE scores in school networks. Finally, the more sophisticated IERMo model for session QoE scores is depicted in green. It can be seen that the IERMo CDFs start off similar to the CDFs of average QoE scores (blue), but generally give slightly higher or slightly lower QoE scores to sessions than the average QoE score. The reason is that the memory effect of the IERMo model is both beneficial and disadvantageous for sessions, in which a video was streamed. The typically high QoE score of a video, cf. Figure 1, results in a positive or negative update of a potential subsequent web QoE score, and thus potentially in slightly higher or lower IERMo session QoE score compared to the average session QoE score.

However, since video streaming has higher network require-

ments than browsing webpages, it seems implausible that, within the same session, the video QoE should be significantly higher than the web QoE, as it was observed in this monitoring study. To take a closer look at this phenomenon, for each session in school networks, the average QoE score of all streamed videos and the average QoE score of all browsed websites are compared. The CDF of the resulting differences within the same session is visualized in Figure 3 in blue. It can be seen that 94% of the sessions have a positive difference, which means that video QoE was higher than web QoE in these sessions. If network administrators wanted to balance these differences, additional 0.2 Mbps would have to be allocated for the web pages of each user, see the orange CDF. The green CDF would actually be a more expected distribution where web QoE should generally be higher than video QoE due to lower network requirements of web browsing. However, an additional capacity of 1 Mbps for web browsing would be needed per user to reach this distribution. It is obvious that these additional capacity requirements are highly unrealistic and implausible.

Nevertheless, this analysis showed that there is a huge discrepancy between the used web QoE model and the video QoE model, although both models can be considered standard and/or state-of-the-art models. It could be observed that these QoE models are incompatible, and thus, cannot be reliably used for network management to quantify and improve the QoE of sessions with both web browsing and video streaming. Thus, there is a need to research and develop compatible QoE models in the future, if QoE-aware network management should not remain limited to a single service.

V. CONCLUSION

The novel Internet-supported educational paradigm is based on reliable devices and high quality Internet access. Education Service Providers (ESPs) have a paramount role in this context, as they must guarantee a high QoE for teachers and students to enable a succesful learning experience. In a collaboration with a major Uruguayan ESP, nationwide QoE measurements were collected from more than 800 schools and about 4000 devices. The QoE monitoring was implemented as a browser plugin, which could log user behavior and key performance indicators (KPIs) of web browsing and video streaming.

When analyzing the QoE of web sessions to derive more detailed insights, which can be leveraged for QoE-aware network management, we identified a fundamental issue with the compatibility of QoE models for web browsing and video streaming. It could be observed that these QoE models are incompatible, and thus, cannot be reliably used for network management to quantify and improve the QoE of sessions with both web browsing and video streaming. The identified incompatibility of QoE models for different services inhibits the successful application of QoE-aware network management for multiple services. Thus, there is a need to conduct novel QoE studies to research and develop compatible QoE models in the future to be able to provide QoE-aware network management for multiple services.

REFERENCES

[1] Plan Ceibal, "What is Plan Ceibal," 2020, Accessed: 2020-09-30. [Online]. Available: https://www.ceibal.edu.uy/en/institucional

[2] P. Le Callet, S. Möller, and A. Perkis (eds), "Qualinet White Paper on Definitions of Quality of Experience," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Tech. Rep., 2013, version 1.2.

[3] J. Nielsen, *Usability Engineering*, 1993.

[4] S. Egger, T. Hoßfeld, R. Schatz, and M. Fiedler, "Waiting Times in Quality of Experience for Web Based Services," in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, 2012.

[5] T. Hoßfeld, F. Metzger, and D. Rossi, "Speed Index: Relating the Industrial Standard for User Perceived Web Performance to Web QoE," in *Proceedings of the 10th International Conference on Quality of Multimedia Experience (QoMEX)*, Pula, Italy, 2018.

[6] E. Bocchi, L. De Cicco, and D. Rossi, "Measuring the Quality of Experience of Web Users," *ACM SIGCOMM Computer Communication Review*, vol. 46, no. 4, 2016.

[7] M. Trevisan, I. Drago, and M. Mellia, "PAIN: A Passive Web performance indicator for ISPs," *Computer Networks*, vol. 149, 2019.

[8] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[9] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A Time-varying Subjective Quality Model for Mobile Streaming Videos with Stalling Events," in *Proceedings of SPIE Applications of Digital Image Processing XXXVIII*, San Diego, CA, USA, 2015.

[10] K. Zeng, H. Yeganeh, and Z. Wang, "Quality-of-experience of Streaming Video: Interactions between Presentation Quality and Playback Stalling," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.

[11] M. Seufert, T. Hoßfeld, and C. Sieber, "Impact of Intermediate Layer on Quality of Experience of HTTP Adaptive Streaming," in *Proceedings of the 11th International Conference on Network and Service Management (CNSM)*, Barcelona, Spain, 2015.

[12] International Telecommunication Union, "ITU-T Recommendation P.1203: Parametric Bitstream-based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services over Reliable Transport," 2016, Accessed: 2020-09-30. [Online]. Available: https://www.itu.int/rec/T-REC-P.1203/en

[13] M. Seufert, N. Wehner, and P. Casas, "Studying the Impact of HAS QoE Factors on the Standardized QoE Model P. 1203," in *Proceedings of the 3rd Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE)*, Vienna, Austria, 2018.

[14] StatCounter, "Browser Market Share Uruguay," 2020, Accessed: 2020-09-30. [Online]. Available: https://gs.statcounter.com/browser-market-share/all/uruguay

[15] World Wide Web Consortium (W3C), "Navigation Timing Level 2," 2020, W3C Working Draft – Accessed: 2020-09-30. [Online]. Available: https://www.w3.org/TR/navigation-timing-2/

[16] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YoMoApp: a Tool for Analyzing QoE of YouTube HTTP Adaptive Streaming in Mobile Networks," in *Proceedings of the European Conference on Networks and Communications (EuCNC)*, Paris, France, 2015.

[17] M. Seufert, N. Wehner, F. Wamser, P. Casas, A. D'Alconzo, and P. Tran-Gia, "Unsupervised QoE Field Study for Mobile YouTube Video Streaming with YoMoApp," in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[18] M. Seufert, "Quality of Experience and Access Network Traffic Management of HTTP Adaptive Video Streaming," Doctoral Thesis, University of Würzburg, 2017. [Online]. Available: https://opus.bibliothek.uni-wuerzburg.de/files/15413/Seufert_Michael_Thomas_HTTP.pdf

[19] Web Hypertext Application Technology Working Group (WHATWG), "HTML Living Standard – 7.9 Session history and navigation," 2020, Last Updated 28 September 2020 – Accessed: 2020-09-30. [Online]. Available: https://html.spec.whatwg.org/multipage/history.html

[20] G. Capdehourat, G. Marín, and A. I. Rodríguez, "Plan Ceibal's Wireless Network for the Implementation of the 1:1 Educational Model," in *Proceedings of the Latin America Networking Conference (LANC)*, Montevideo, Uruguay, 2014.

[21] G. Capdehourat, C. Aguerrebere, F. Bascans, G. Álvarez, and P. Porteiro, "Unveiling wi-fi usage profiles: A first look at a K-12 education service provider," in *IEEE INFOCOM Workshops*, 2019.

[22] L. D. Catledge and J. E. Pitkow, "Characterizing Browsing Strategies in the World-Wide Web," *Computer Networks and ISDN Systems*, vol. 27, no. 6, pp. 1065–1073, 1995.

[23] A. Bianco, G. Mardente, M. Mellia, M. Munafò, and L. Muscariello, "Web User-session Inference by means of Clustering Techniques," *IEEE/ACM Transactions On Networking*, vol. 17, no. 2, pp. 405–416, 2008.

[24] L. Vassio, I. Drago, M. Mellia, Z. B. Houidi, and M. L. Lamali, "You, the Web, and Your Device: Longitudinal Characterization of Browsing Habits," *ACM Transactions on the Web (TWEB)*, vol. 12, no. 4, pp. 1–30, 2018.

[25] T. Hoßfeld, S. Biedermann, R. Schatz, A. Platzer, S. Egger, and M. Fiedler, "The Memory Effect and its Implications on Web QoE Modeling," in *Proceedings of the 23rd International Teletraffic Congress (ITC)*, San Francisco, CA, USA, 2011.