Beauty is in the Eye of the Smartphone Holder A Data Driven Analysis of YouTube Mobile QoE

Nikolas Wehner*, Sarah Wassermann‡, Pedro Casas†, Michael Seufert† and Florian Wamser*
* University of Würzburg, † AIT Austrian Institute of Technology, ‡ Inria Paris
{nikolas.wehner.fl, pedro.casas, michael.seufert.fl}@ait.ac.at, sarah.wassermann@inria.fr, florian.wamser@informatik.uni-wuerzburg.de

***Abstract*—Measuring the Quality of Experience (QoE) undergone by cellular network users has become paramount for cellular ISPs. Given its overwhelming dominance and ever-growing popularity, this paper focuses on the analysis of QoE for YouTube in mobile networks. Using a large-scale dataset of crowdsourced YouTube QoE measurements collected in smartphones with YoMoApp, we analyze the evolution of multiple relevant QoE-related metrics over time for YouTube mobile users. The dataset includes measurements from more than 360 users worldwide, spanning over the last five years. Our data-driven analysis shows a systematic performance and QoE improvement of YouTube in mobile devices over time, accompanied by an improvement of cellular network performance and by an optimization of the YouTube streaming behavior for smartphones.**

***Index Terms*—YouTube; Mobile Network Measurements; Quality of Experience; Crowdsourcing.**

I. Introduction

Smartphones are today the most popular Internet access means, and the tendency is towards even higher dominance and adoption. While this growth has a direct economical benefit for cellular Internet Service Providers (ISPs), it also posses complex challenges in terms of cellular network management and operation. One of these challenges relates to the monitoring and assessment of the network's performance as perceived by the end users. In this paper we study the Quality of Experience (QoE) of mobile networks and services in a completely data-driven manner, by relying on a large-scale dataset of QoE measurements passively collected at users' smartphones. We particularly focus on the YouTube video steaming service, based on its popularity and dominance in terms of traffic volume within mobile networks.

Measurements are collected with YoMoApp [1], [2], an app for crowsourced YouTube QoE measurements publicly available at the Google Play Store. The dataset we study consists of more than 3000 YouTube video sessions collected worldwide over 70 different cellular ISPs and from more than 360 different users, from 2014 till today. YoMoApp measures a very rich set of QoE-related and QoS key performance indicators (KPIs) of YouTube mobile on different layers of the communications' stack, from application-layer KPIs such as re-buffering events and quality changes to network-layer KPIs such as transmitted bytes, throughput, radio access technology (RAT), handovers, etc. The app also collects data about the device and the user context, including KPIs such as screen size and orientation, user location and mobility, ISP, etc. Finally, it also collects user feedback in terms of user experience and satisfaction immediately after a video session has ended.

Through the analysis of the measurements we are able to observe quite interesting results in terms of YouTube QoE evolution on the long run. More precisely, we observe a systematic performance and QoE improvement of YouTube in mobile devices over time, accompanied by an (expected) improvement of cellular networks' performance and by an optimization of the YouTube streaming behavior for smartphones. Our study shows that these improvements have a direct impact on the user engagement in YouTube mobile, with an increase of more than 30% on the relative video consumption time.

The remainder of the paper is organized as follows: Sec. II briefly overviews the related work, focusing on the specific case of YouTube mobile QoE analysis. Sec. III describes the YoMoApp application, and reports the findings obtained from the analysis of five years of YoMoApp measurements collected between 2014 and 2018, focusing specially on the evolution of QoE-related and network performance related metrics.

II. Related Work

There is a long literature covering QoE-based network monitoring approaches, most of them focused on fixed-line networks or relying on in-network measurements. A good summary describing such approaches and the corresponding challenges around this topic is provided in [3]. When it comes to the specific case of video streaming QoE monitoring, we found multiple approaches mapping either in-network or in-device/application measurements to QoE metrics. In an early work [4] authors propose YoMo, a client-side Deep Packet Inspection (DPI)-based and application-based tool to monitor YouTube video flows and buffered playtime at the video player side, from where playback stallings are derived. Multiple subsequent papers [5], [6], [7] extended the YoMo approach to perform YouTube QoE monitoring at the ISP-scale - both fixed and mobile networks, relying on DPI-based techniques. Some more recent papers [8], [9], [10] also adopted browser plug-in-based measurements to passively monitor video QoE-relevant KPIs such as initial delay, stalling, and quality switches. The advantage of application-level monitoring is that most QoE-relevant information can be accessed directly and accurately, and does not need to be estimated.

Focusing exclusively on mobile networks and devices, there is an assorted list of tools to measure QoE-based network

978-3-903176-14-0 © 2018 IFIP

performance: some examples include Mobilyzer [11] and Netalyzr [12]. QoE Doctor [13] measures mobile app QoE, using active measurements at both application and network layers. Similar tools for YouTube measurement at smartphones are presented in [14], [15]. In [1], [2] we introduced YoMoApp, an app to passively monitor YouTube QoE-related features in smartphones, extending the original YoMo paper [4]. Previous work has also focused on the usage of machine learning techniques to predict QoE for mobile apps: for example, authors in [16] and ourselves in [17], [18] propose machine-learning based approaches to evaluate mobile apps QoE using passive in-network and/or in-device measurements. Similarly, papers such as [19], [20], [21] also focus on machine learning models to infer or predict QoE-relevant metrics.

III. YouTube Mobile QoE Analysis

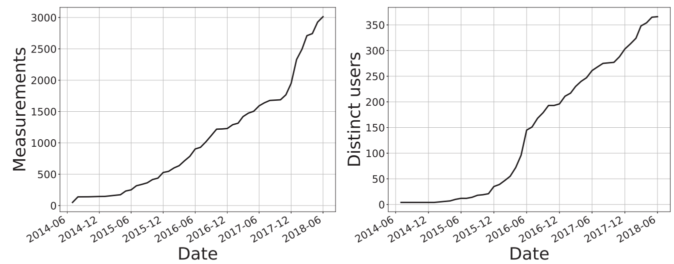
In this section we analyze the complete dataset of measurements collected by YoMoApp over the past five years. Using these measurements, we characterize the temporal evolution of YouTube mobile in terms of QoE-relevant metrics, user behavior as well as cellular network performance.

A. YouTube Mobile Analysis with YoMoApp

The goal of YoMoApp [1], [2] is to provide a distributed, crowdsourced-based monitoring platform to monitor QoE user feedback and application layer KPIs of YouTube mobile that have a high correlation with the actual QoE of the YouTube mobile app users. KPIs such as initial delay, stallings, and quality adaptation are the most relevant QoE-related features measured by YoMoApp. These are passively collected during the playback of a video session, from the state and buffer of the video player, as well as from the resolution of the played-out video. Measurements collected at each device are locally logged and periodically exported to a cloud server, which can then be accessed and further analyzed through the YoMoApp cloud-dashboard[1].

Besides the monitoring of the playback, network and context parameters are also collected by YoMoApp. Several device characteristics and their changes, namely, screen size, screen orientation, volume, player size, and player mode (normal/full screen) are monitored. Network usage is also monitored. The amount of download and upload bytes for the device, for the mobile network, and only for YoMoApp are polled every second. Moreover, changes of operator, RAT, cell ID, signal strength, and GPS position are also collected. YoMoApp also collects the user QoE feedback when a video session has ended or is aborted. The user is asked to assess the QoE of the session on a 5-point ACR MOS scale ranging from 1 (bad) to 5 (excellent) [22] through different questions. The different feedback ratings requested to the user include his opinion on the video quality, the streaming quality, the video itself, and the acceptability of the service - the latter is a binary feedback. The feedback is requested only if the user wishes to provide it, which can be decided at the time of starting the app.

[1]`http://yomoapp.de/dashboard`



(a) Cumulative num. of sessions. (b) Cumulative num. of users.

Figure 1: Number of sessions and distinct users over time.

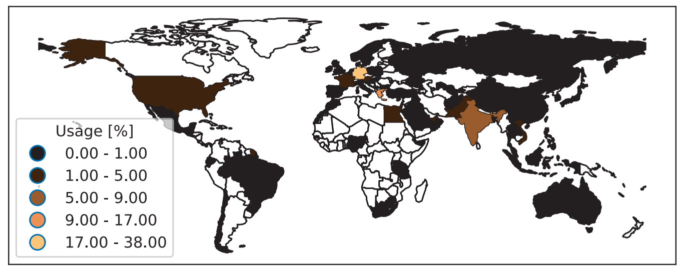


Figure 2: Worldwide usage of YoMoApp.

The full list of collected KPIs for the individual log files are summarized in Tab. I. In both the `data` log file and the `events` log file, measurements are synchronized with a corresponding Unix timestamp. The monitoring of these KPIs is performed either every second, e.g., the monitoring of the current playtime, or only when an event occurs, e.g., a change in the played out video quality. In contrast, the `stats` log file offers an overview/aggregation of the video streaming session. Full details on the collected measurements are available at the YoMoApp documentation[2].

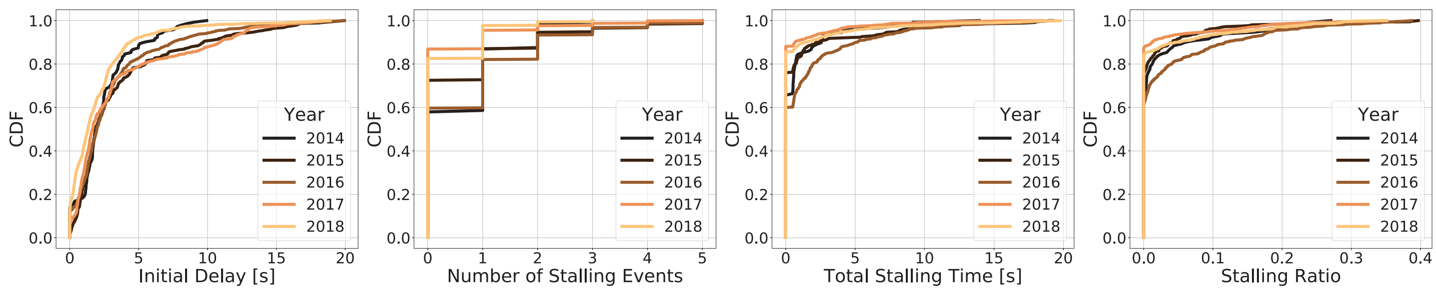
B. YouTube Data-Driven QoE Analysis

We now analyze the complete dataset of measurements collected by YoMoApp over the past five years. Using these measurements, we characterize the temporal evolution of YouTube mobile in terms of QoE-relevant metrics, user behavior as well as cellular network performance. The complete dataset contains today 3013 valid streaming sessions, which were monitored in the period from July 2014 to June 2018. These sessions originate from 366 different users from all over the world. Fig. 1a depicts the cumulative number of video sessions streamed by YoMoApp over time, and Fig. 1b shows the same evolution for the cumulative number of distinct users (devices). There is a clear trend in the number of new users from beginning 2016 onwards, which is linked to a more aggressive dissemination and advertisement approach to promote the usage of YoMoApp around different communities. Since January 2017 the number of video sessions, collected measurements and new users has more than doubled. Particularly interesting is the fact that during the first half of 2018 we could observed more than 900 new sessions, which largely exceeds the total

[2]`http://yomoapp.de/documentation.pdf`

Log file	Parameters				
data	Current playtime	Buffer	Available playtime		
events	Video-ID	Quality	Network	Received bytes	Transmitted bytes
	Cell-ID	Signal	SSID	BSSID	RSSI
	Location	Title	Duration	Screen orientation	Player size
	Player mode	Volume	MSE	Supported codecs	Player state
	Dialog	Content rating	Quality rating	Streaming rating	Acceptability rating
	YouTube loading time	Advertisement	Video end	App behavior	Hyperlink
stats	Date	Time	Device-ID	Mobile operator	Country
	Network switches	Networks	Screen size	Screen density	Orientation changes
	Orientations	Player resizes	Player sizes	Handovers	Cell-ID
	Video-ID	Video title	Log time	Length of video	User engagement
	Initial delay	Quality switches	Qualities	Stalling events	Total stalling time
	Average stalling time	Maximal stalling time	Average buffer	Maximal buffer	Pause events
	Content rating	Quality rating	Streaming rating	Acceptability rating	

Table I: Monitored KPIs per log file in YoMoApp.

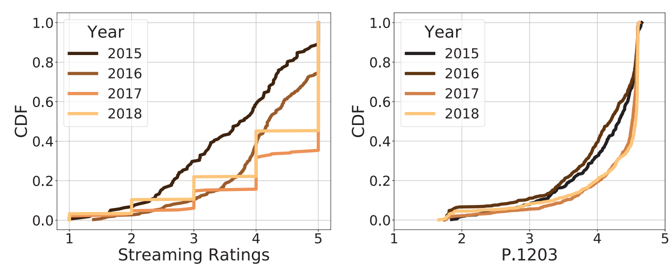


(a) initial delay (b) number of stallings (c) total stalling time (d) stalling ratio

Figure 3: Temporal evolution of the performance of YouTube mobile streaming in terms of QoE-relevant KPIs.

number of sessions monitored in 2017. This indicates a very positive usage trend.

The worldwide distribution of the collected YoMoApp measurements is shown in Fig. 2 as a heatmap-like diagram. Most measurements were collected in Germany (38%), Greece (17%), India (9%), and France (5%). Other participating countries show a share of equal to or less than 3%. Overall, we have collected measurements distributed over 58 different countries. It is important to remark that all YoMoApp users perform the measurements on their own devices and cellular ISPs, which results in a very rich and diverse dataset.



(a) Subjective ratings. (b) P.1203 scores.

Figure 4: Distribution of MOS scores per session.

YouTube QoE has improved over time: we move on now to the analysis of multiple QoE-relevant KPIs for YouTube video streaming. In particular, we focus on the temporal evolution of the initial playback delay, number of stalling events, total stalling time and re-buffering or stalling ratio. Fig. 3 reports the distribution of these four metrics, split by year. The first interesting observation is that, excluding 2014 and 2015 which had a smaller number of sessions, one can clearly appreciate an improvement over time on all the QoE-relevant metrics, with 2018 sessions showing the smallest initial delays and best performance in terms of less stalling events. As of 2018, more than 90% of the video sessions experience an initial playback delay below 5 seconds, and almost 90% of the sessions playout smoothly without re-buffering events. In contrast, the initial delay for video sessions in 2016 was below 5 seconds for 80% of the sessions, and only 60% of the 2016 sessions experienced no stallings. When considering highly QoE-impaired video sessions, we see that more than 12% of the video sessions in 2016 had a re-buffering ratio above 10%, whereas this number reduces to about 5% in 2017 and 2018.

Fig. 4 reports the distribution of (a) the actual subjective QoE feedback reported by the users and (b) an estimation of the QoE by applying the standardized ITU-T P.1203 model [23]. We particularly focus on the feedback reported by the users in terms of their quality perception for the video streaming. As we discussed before, there is a clear improvement in the QoE as reported by the users in the last couple of years, with about 80% of the sessions being rated as very good or excellent (MOS 4 or 5), in contract to the 40% to 60% reported in past years. As shown in Fig. 4b,

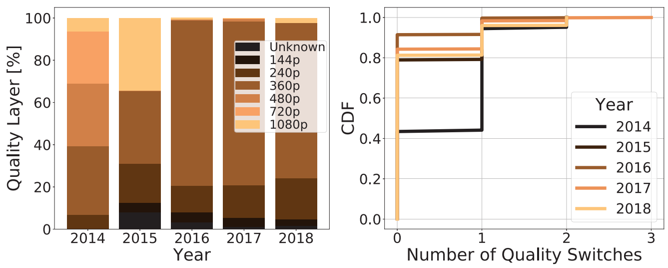
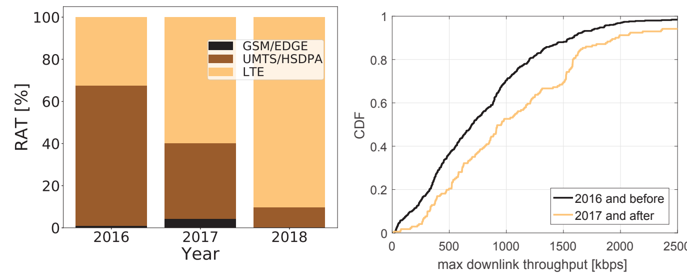


Figure 5: Video quality levels and quality switches.



(a) Radio access technology.
(b) Max. downlink throughput.

Figure 6: Radio access and video download throughput.

these results are quite accurately captured by the predictions obtained through the P.1203 model.

YouTube mobile video distribution is today more efficient than in the past: the played out video quality levels grouped by year and the distribution of the number of quality switches per year are displayed in Fig. 5. The distribution of requested video qualities by the YoMoApp video player reveals that in contrast to the period from 2016 to 2018, back in 2014 and 2015 the played out video qualities varied much stronger, with a higher prevalence of higher quality levels as compared to today. The YouTube streaming service has been evolving over time, not only for the fixed-line network scenario, but mainly in mobile networks. When YouTube started playing in mobile devices, the adaptive streaming policy was less conservative and higher quality levels would be requested in adaptive streaming mode. From 2016 onwards, the most dominant video quality changed to 360p, which is a more conservative quality level, imposing less bandwidth requirements. There are also videos with lower video qualities like 240p, but almost no HD content was streamed within the last three years with YoMoApp. This is perfectly aligned to our previous findings on YouTube QoE in smartphones [24], where we observed that lower resolution results in the same subjective experience as higher resolutions when dealing with smartphones, due to the small screen-sizes. Thus, it makes little sense and is less efficient to stream HD contents to smartphones.

As a consequence, it is also not surprising that the number of quality switches observed within the last three years is much lower compared to 2014 and 2015. Fig. 5b shows that no quality switch could be observed for more than 80% of the sessions in the period 2016 to 2018, meaning that the initial quality selected by YouTube mobile was matching the underlying network performance. In contrast, in 2014 only 43% of the sessions showed no quality switch, around 53% observed one quality switch, and the remaining sessions resulted in two or more quality switches.

Mobile network technology and performance have also improved, potentially resulting in increased user engagement: the distribution of the underlying RAT per year is displayed in Fig. 6a. We differentiate between 2G (GSM/EDGE), 3G (UMTS/HSDPA) and 4G (LTE). RAT information started being collected only from 2016 on. In 2016, UMTS/HSDPA was the dominant RAT, with a prevalence of about 66% of all sessions with cellular access. In 2017, the balance shifted and LTE became the dominant RAT with a share of 59%. This dominance increased even more in 2018, where sessions with LTE make up to 90% of all streaming sessions with cellular access. As a consequence, better network performance is observed over time. For example, Fig. 6b shows the distribution of the maximum download throughput achieved by YoMoApp video sessions before and after December 2016. The average max. download throughput increased from about 2Mbps to more than 10Mbps, and the median has also increased from about 600kbps to 1Mbps.

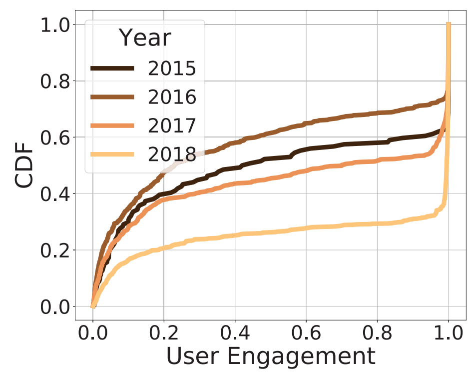


Figure 7: Evolution of user engagement.

Finally, the user engagement distribution per year is depicted in Fig. 7. User engagement defines the fraction of the total video length a user watched, before the video was aborted or the video ended (100% user engagement). It started being measured in 2015, thus we have no results for 2014. Results show how user engagement has systematically increased over time, and significantly in 2018. More than 60% of the videos were watched completely and only 20% of the users aborted the video at 20% or less of the video playback. This indicates that YoMoApp is increasingly being used as a standard video player. The increased user engagement can also be explained by the improvement of the network performance in terms of higher downlink throughputs, as well as by the enhanced QoE.

We do hope that the full YoMoApp distributed monitoring system (i.e, app and dashboard) would bring many interesting opportunities for researchers, industry players and/or interested end-users alike.

REFERENCES

[1] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "Understanding YouTube QoE in Cellular Networks with YoMoApp – a QoE Monitoring Tool for YouTube Mobile," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom)*, Paris, France, 2015.

[2] ——, "Yomoapp: A tool for analyzing qoe of youtube http adaptive streaming in mobile networks," in *2015 European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 239–243.

[3] W. Robitza, A. Ahmad, P. A. Kara, L. Atzori, M. G. Martini, A. Raake, and L. Sun, "Challenges of future multimedia qoe monitoring for internet service providers," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22 243–22 266, Nov 2017.

[4] B. Staehle, M. Hirth, R. Pries, F. Wamser, and D. Staehle, "YoMo: A YouTube Application Comfort Monitoring Tool," in *Proceedings of the 1st Workshop of Quality of Experience for Multimedia Content Sharing (QoEMCS)*, Tampere, Finland, 2010.

[5] R. Schatz, T. Hoßfeld, and P. Casas, "Passive YouTube QoE Monitoring for ISPs," in *Proceedings of the 2nd International Workshop on Future Internet and Next Generation Networks (FINGNet)*, Palermo, Italy, 2012.

[6] P. Casas, R. Schatz, and T. Hoßfeld, "Monitoring YouTube QoE: Is Your Mobile Network Delivering the Right Experience to Your Customers?" in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, 2013.

[7] P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 2, pp. 44–46, 2013.

[8] H. Nam, K.-H. Kim, D. Calin, and H. Schulzrinne, "YouSlow: A Performance Analysis Tool for Adaptive Bitrate Video Streaming," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 111–112, 2014.

[9] H. Nam, K.-H. Kim, and H. Schulzrinne, "QoE Matters More Than QoS: Why People Stop Watching Cat Videos," in *Proceedings of the IEEE INFOCOM*, San Francisco, CA, USA, 2016.

[10] H. Nam and H. Schulzrinne, "YouSlow: What Influences User Abandonment Behavior for Internet Video?" Columbia University, Tech. Rep., 2016.

[11] A. Nikravesh, H. Yao, S. Xu, D. Choffnes, and Z. M. Mao, "Mobilyzer: An open platform for controllable mobile network measurements," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '15. New York, NY, USA: ACM, 2015, pp. 389–404.

[12] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzr: Illuminating the edge network," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 246–259. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879173

[13] Q. A. Chen, H. Luo, S. Rosen, Z. M. Mao, K. Iyer, J. Hui, K. Sontineni, and K. Lau, "Qoe doctor: Diagnosing mobile app qoe with automated ui control and cross-layer analysis," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14. New York, NY, USA: ACM, 2014, pp. 151–164. [Online]. Available: http://doi.acm.org/10.1145/2663716.2663726

[14] I. Ketykó, K. De Moor, T. De Pessemier, A. J. Verdejo, K. Vanhecke, W. Joseph, L. Martens, and L. De Marez, "QoE Measurement of Mobile YouTube Video Streaming," in *Proceedings of the 3rd Workshop on Mobile Video Delivery (MoViD)*, Florence, Italy, 2010.

[15] G. Gómez, L. Hortigüela, Q. Pérez, J. Lorca, R. García, and M. C. Aguayo-Torres, "YouTube QoE Evaluation Tool for Android Wireless Terminals," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 164, pp. 1–14, 2014.

[16] V. Aggarwal, E. Halepovic, J. Pang, S. Venkataraman, and H. Yan, "Prometheus: Toward Quality-of-Experience Estimation for Mobile Apps from Passive Network Measurements," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile)*, Santa Barbara, CA, USA, 2014.

[17] P. Casas, A. D'Alconzo, F. Wamser, M. Seufert, B. Gardlo, A. Schwind, P. Tran-Gia, and R. Schatz, "Predicting qoe in cellular networks using machine learning and in-smartphone measurements," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–6.

[18] P. Casas, M. Seufert, F. Wamser, B. Gardlo, A. Sackl, and R. Schatz, "Next to you: Monitoring quality of experience in cellular networks from the end-devices," *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, pp. 181–196, June 2016.

[19] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki, "Measuring Video QoE from Encrypted Traffic," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, Santa Monica, CA, USA, 2016.

[20] I. Orsolic, D. Pevec, M. Suznjevic, and L. Skorin-Kapov, "A machine learning approach to classifying youtube qoe based on encrypted network traffic," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22 267–22 301, Nov 2017.

[21] M. H. Mazhar and M. Z. Shafiq, "Real-time video quality of experience monitoring for https and quic," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018.

[22] International Telecommunication Union, "ITU-T Recommendation P.800: Methods for Subjective Determination of Transmission Quality," 1996.

[23] ——, "ITU-T Recommendation P.1203: Parametric Bitstream-based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services over Reliable Transport," 2016. [Online]. Available: https://www.itu.int/rec/T-REC-P.1203/en

[24] P. Casas, R. Schatz, F. Wamser, M. Seufert, and R. Irmer, "Exploring qoe in cellular networks: How much bandwidth do you need for popular smartphone apps?" in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, ser. AllThingsCellular '15. New York, NY, USA: ACM, 2015, pp. 13–18. [Online]. Available: http://doi.acm.org/10.1145/2785971.2785978