On the Analysis of QoE in Cellular Networks: from Subjective Tests to Large-scale Traffic Measurements

Pedro Casas*, Martin Varela†, Pierdomenico Fiadino*, Mirko Schiavone*, Helena Rivas†, Raimund Schatz*

*FTW – Telecommunications Research Center Vienna
{surname}@ftw.at

†VTT Technical Research Center of Finland
{name.surname}@vtt.fi

***Abstract*—Mobile devices such as smartphones are taking over traditional devices for Internet access in today's scenario, and the near future forecast is overwhelming: by 2016, a quarter of the world population will be using smartphones to access the Internet. In this context, understanding the Quality of Experience (QoE) of popular services in mobile devices becomes paramount for cellular network operators, who need to offer high quality levels to reduce the risks of customers churning for quality dissatisfaction. In this paper we study the problem of QoE provisioning in mobile devices, presenting the results obtained from subjective lab tests performed for popular end-user services accessed through smartphones. Our analysis addresses the impact of access downlink bandwidth on the QoE of four different popular services and mobile apps: Facebook, Web browsing through Chrome, Google Maps, and WhatsApp. The study also considers the characterization of WhatsApp QoE in a real setting, mapping the lab results to large-scale measurements conducted in a major cellular network. The results presented in this paper provide a sound basis for better understanding the QoE requirements of popular services and mobile apps, as well as for dimensioning the underlying provisioning network. To the best of our knowledge, this is the first paper combining QoE lab-test results for mobile devices with large-scale measurements in an operational cellular network.**

Keywords*—QoE; Mobile Networks; Smartphones; Subjective Lab Tests; Large-scale Measurements.

I. Introduction

Smartphones are becoming the most typical type of device to access Internet today. Recent projections [2] show that by 2016, a quarter of the world population will be using smartphones to access the most popular services such as YouTube, Facebook and WhatsApp. According to Cisco's global mobile data traffic forecast [1], smartphones will be responsible for more than three-quarters of the mobile data traffic generated by 2019. In the light of these trends, cellular network operators are becoming more and more interested in understanding how to dimension their access networks and how to manage their customers' traffic to capture as many new customers as possible. In this scenario, the concept of Quality of Experience (QoE) has the potential to become one of the main guiding paradigms for managing quality in cellular networks. Closely linked to the subjective perception of the end-user, QoE enables a broader, more holistic understanding of the factors that influence the performance of systems, complementing traditional technology-centric concepts such as Quality of Service (QoS). Indeed, QoE is today an important differentiator between providers, but most of the times, operators do not really grasp the key aspects related to QoE in their networks.

In this paper we claim that understanding QoE in mobile devices is paramount for cellular network operators, and present the results obtained from subjective lab tests performed for popular end-user services accessed through smartphones. In particular, we consider the following well-known applications in mobile devices: Facebook, Google Maps (Gmaps from now on), Web browsing through Google Chrome, and WhatsApp. The evaluations performed in these subjective tests consider the impact of downlink bandwidth on end-user QoE, which represents one of the most relevant QoS-based characteristics of the access network. The main contribution of our study is to shed light into the problem of QoE-based network provisioning for mobile devices, offering a comprehensive analysis of the QoE undergone by users when the underlying access network presents different QoS characteristics or performance levels.

The standard approach to assess the performance of networks and services from a QoE end-user perspective is to conduct controlled lab experiments [16]–[18]. The key benefits of such an approach rely on the full control the experimenter has on the overall evaluation process. Indeed, content and context are fully known and controlled, and users are directly briefed and observed on the spot, providing as such tangible and solid results.

However, lab experiments are small scale by nature, and lack as such the intrinsic characteristics of a more realistic scenario, in which a large population of thousands of users interact with the tested services. The statistical properties of the traffic generated by such large populations are generally different, thus mapping the QoE results of lab studies to the real field becomes paramount to better understand the real implications behind the obtained QoS/QoE mappings. For this reason, we take a step further in this direction and present a characterization of the QoE of one of the tested services "in the wild", by mapping the obtained results from the lab tests into large-scale measurements conducted on the cellular network of a nationwide Europen operator. In particular, we focus on the paradigmatic case of WhatsApp, given its high popularity and evergrowing adoption by users with mobile devices.

The remainder of the paper is organized as follows: Section II presents an overview of the related work on QoE for web and cloud-based services, focusing on the specific case of mobile

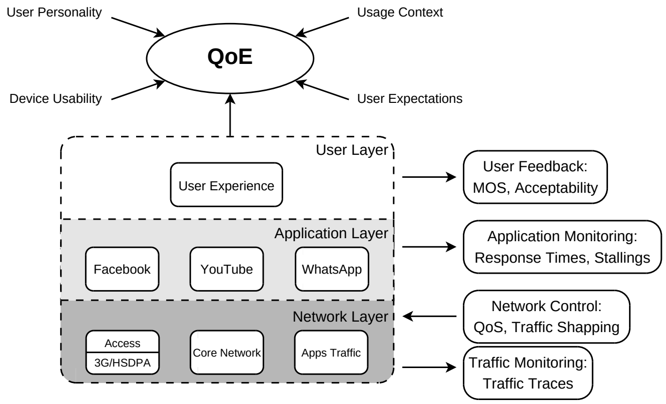


Figure 1. Layered QoE evaluation methodology for networking services.

devices. Section III describes the evaluation methodology and the experimental setup used in the QoE subjective tests. Section IV presents the main results of the subjective tests, covering the impact of the downlink bandwidth on the overall experience and acceptability of the end-user when accessing the considered applications. The particular characterization of QoE in WhatsApp through large-scale measurements is discussed in Section V. Finally, Section VI concludes this work.

II. Related Work

The study of the QoE requirements for web-based services and cloud-based applications as the ones we target in this paper has a long list of fresh and recent references. A good survey of the QoE-based performance of mobile networks when accessing many different web and cloud services is presented in [4]. Studies on the QoE of popular services such as Facebook [3], cloud storage [6], remote virtual desktop [7], and less popular but more trendy such as cloud gaming [5] offer a good source of information to understand how to manage the network for such services. The main limitation of these papers when considering the analysis of performance in cellular networks is that the considered access devices are not smartphones, but rather traditional laptops with mobile broadband connections. Standardization efforts have also been devoted to the analysis of user-based, end-to-end performance of data applications [19], particularly targeting web-based services.

When it comes to our specific analysis of QoE in mobile networks and mobile devices, references become scarcer, showing that there is still an important gap to fill. Authors in [13] describe a subjective QoE evaluation framework for mobile Android devices in a lab environment. In [14], authors study the QoE of YouTube in mobile devices through a field trial. Authors in [15] recently introduced Prometheus, an approach to estimate QoE of mobile apps, using both passive in-network measurements and in-device measurements, applying machine learning techniques to obtain mappings between QoS and QoE.

Finally, WhatsApp is a very new service and its study has been so far quite limited. Some recent papers have partially addressed the characterization of its traffic, including a QoE perspective [12].

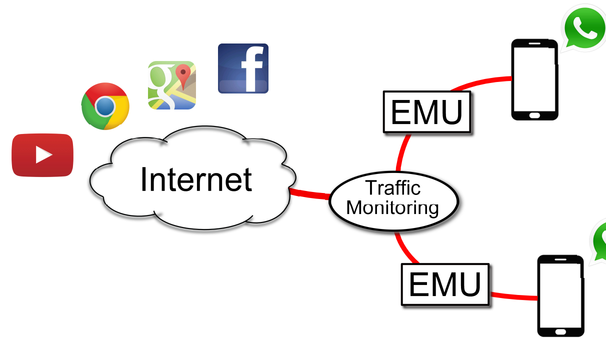


Figure 2. Experimental setup used in the study. Devices are connected to the Internet through independent, controlled WiFi connections.

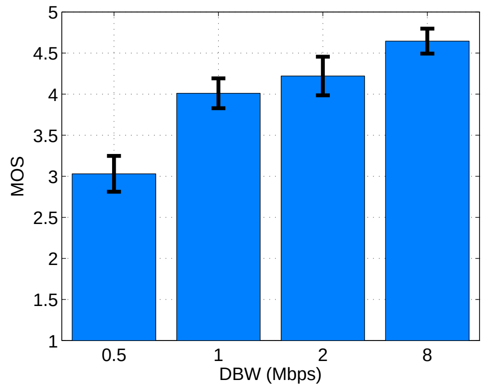
Table I. Operational expected downlink bandwidth values for different access technologies.

Access Technology	Downlink Bandwidth
LTE	30 Mbps
HSPA+	15 Mbps
HSPA	Mbps
UMTS	384 kbps
EDGE	160 kbps
GPRS	40 kbps

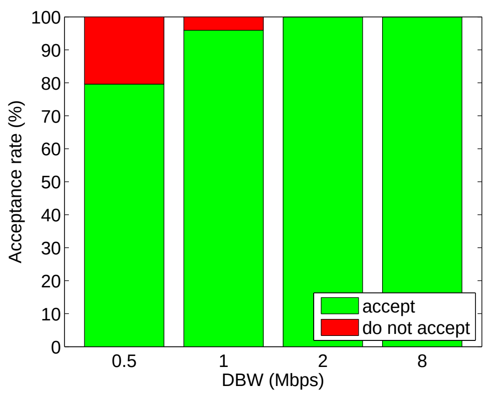
III. Experimental Methodology

Lab tests are realized through the layered evaluation methodology depicted in Figure 1. The experience of a user with any application is conditioned by multiple features, including dimensions such as technical characteristics of the application, user personality and expectations, user demographics, device usability, and usage context among others. Particularly when evaluating networking-based applications, the influence of the network itself as well as its interplay with the particular application have to be linked to the user's opinions, additionally identifying those perceivable performance parameters that are most relevant to the user experience. This mapping is realized by analyzing and correlating the three layers depicted in Figure 1: the *network layer* accounts for the influence of the network QoS parameters (e.g., network bandwidth, RTT, etc.); the *application layer* considers both the technical characteristics (e.g., screen resolution, video bitrate, web-page complexity) and the perceivable performance parameters of the application (e.g., page-load times, response time, video stalling, etc.); finally, the *user layer* spans the user subjective opinions on the evaluated application (e.g., MOS values, acceptability, etc.). The experimental evaluations reported in this work are designed in such a way that all the three aforementioned layers could be properly measured, at least partially in the case of application layer measurements.

The subjective study consists of 52 participants interacting with the aforementioned services while experiencing different bandwidth and access RTT profiles in the background data connection. Figure 2 depicts a high-level diagram of the experimental testbed employed in the subjective tests. Android smartphone devices are used in the study (Samsung Galaxy S4, OS Android 4.4 KitKat). Devices are connected to the

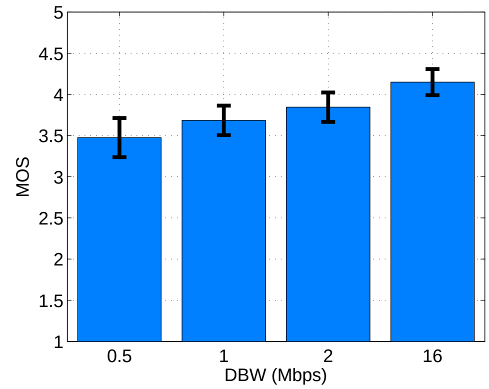


(a) Overall quality (MOS).

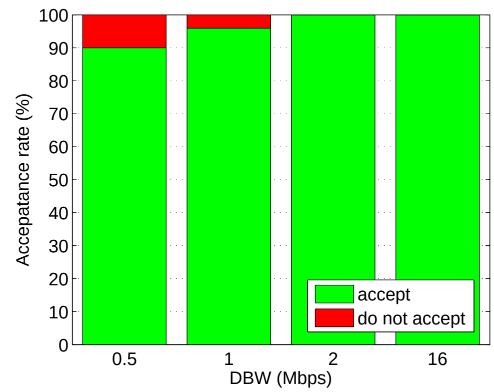


(b) Acceptance rate.

Figure 3. QoE in Facebook. Overall quality and acceptability for different downlink bandwidth configurations. A DBW of 500 kbps is not high enough to reach full user satisfaction, but a DBW of 1 Mbps results in good overall quality, with almost full acceptance.



(a) Overall quality (MOS).



(b) Acceptance rate.

Figure 4. QoE in Web browsing (news website). Overall quality and acceptability for different downlink bandwidth configurations. Overall quality increases in a logarithmic fashion with increasing values of the DBW.

Internet through separate WiFi access networks. The downlink traffic between the different evaluated services and the devices is routed through a modified version of the very well known NetEm network emulator [20] so as to control the different access network profiles under evaluation.

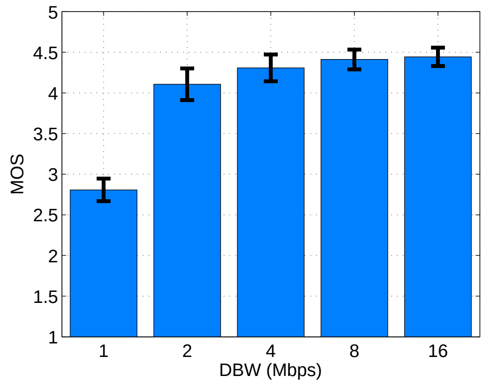
Different constant downlink bandwidth profiles are instantiated at the network emulators, ranging from 0.5 Mbps to 16 Mbps. These bandwidth profiles are selected from operational experience, particularly following typical operational values reported in Table I for different access network technologies (LTE, 3G/2G, etc.). This list is also complemented with operational knowledge coming from cellular operators, collaborating with the project which drives this work, the ACE project[1]. An important remark is that access RTT is kept at 10 ms when downlink bandwidth is varied, which corresponds to near optimal performance in mobile networks (e.g., LTE).

Participants were instructed to perform independent tasks for each of the considered applications. In the case of Facebook, participants were instructed to access the application with a specific user account, browse the timeline of this user, and browse through specific photo albums created for this user. In the WhatsApp tests, participants worked in couples and exchanged specific video files of fixed size (i.e., 5 MB), and the participant downloading the video file was the one providing a QoE evaluation, based on the experienced time. Web browsing tasks consisted of reading and browsing through a popular and complex News website (http://edition.cnn.com/). Finally, Gmaps tasks consisted of exploring different city maps using the Gmaps application, in satellite view, which consumes more bandwidth.

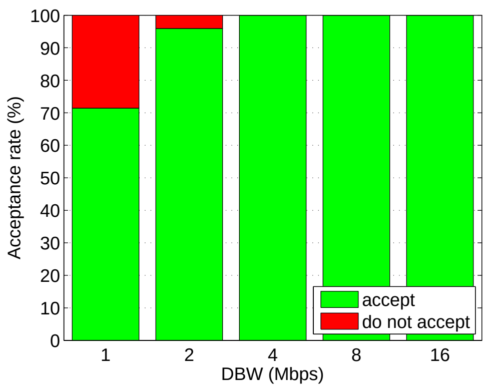
Tests were performed in a dedicated lab for subjective studies, compliant with the QoE subjective studies standards [16]–[18]. All traffic flows are captured and exported to standard `pcap` traces for off-line traffic analysis, using high-performance Endance DAG cards. Regarding participants' demographics, 29 participants were female and 23 male, the average age was 32 years old, with 40 participants being less than 30 years old. Around half of the participants were students and almost 43% were employees, and 70% of the participants have completed university or baccalaureate studies.

Regarding QoE feedback, participants were instructed to rate their *overall experience* (rate the overall quality) according to a continuous ACR Mean Opinion Score (MOS) scale [16], ranging from "bad" (i.e., MOS = 1) to "excellent" (i.e., MOS = 5). MOS ratings were issued by participants through a custom questionnaire application running on separate laptops, which pops up immediately after a condition was tested. Participants also provided feedback on the *acceptability* of the application, stating whether they would continue using the

[1]The ACE QoE project at FTW Vienna, http://ace.ftw.at/

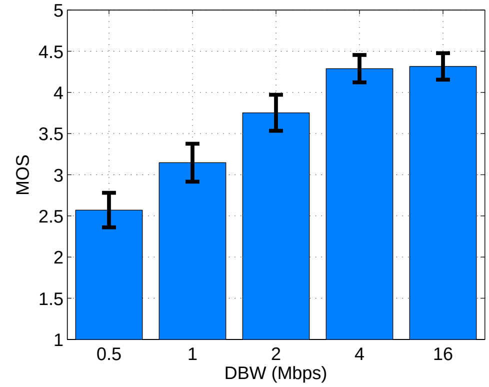


(a) Overall quality (MOS).

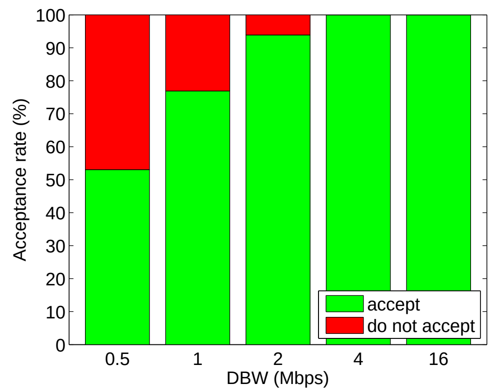


(b) Acceptance rate.

Figure 5. QoE in Gmaps. Overall quality and acceptability for different downlink bandwidth configurations. A DBW of 4 Mbps results in near optimal QoE, and from this value on, QoE saturation already occurs.



(a) Overall quality (MOS).



(b) Acceptance rate.

Figure 6. QoE in WhatsApp. Overall quality and acceptability for different downlink bandwidth configurations. Users tolerate WhatsApp downloads with a good overall experience and high acceptability as long as the DBW is above 2 Mbps, whereas bad quality is observed for a DBW of 500 kbps.

application under the corresponding conditions or not. Each testing session runs for a total time of two hours. Participants were compensated with vouchers for their participation, which proved to be sufficient for achieving correct involvement in the tasks.

IV. QoE in Mobile Devices

In this section we present and discuss the results obtained in the conducted tests. Constant Downlink BandWidth (DBW) profiles are tested for the studied services. Facebook is tested with DBW = 0.5 Mbps, 1 Mbps, 2 Mbps, and 8 Mbps. The profiles for Web browsing are almost identical to those used in Facebook, expect that the last condition corresponds to an optimal DBW = 16 Mbps. Gmaps is tested with a fully logarithmic scale: 1 Mbps, 2 Mbps, 4 Mbps, 8 Mbps, and 16 Mbps. Finally, the WhatsApp DBW profile takes the values 0.5 Mbps, 1 Mbps, 2 Mbps, 4 Mbps, and jumps to 16 Mbps to verify the occurrence of QoE-saturation, which we shall explain next.

A final remark regarding interpretation of results: the reader shall note that the maximum MOS ratings declared by the participants are never 5 but somewhere between 4.2 and 4.6. This is a well known phenomenon in QoE studies called *rating scale saturation*, where users hardly employ the limit values of the scale for their ratings [4]. So from now on, we shall consider as optimal quality a MOS score close to 4.5.

A. *QoE in Facebook Mobile*

Figure 3 reports the results obtained in the Facebook tests for different DBW configurations, considering both (a) the overall quality and (b) the acceptance rate. A DBW of 500 kbps is not high enough to reach full user satisfaction in Facebook mobile for Android devices, as participants declared a fair quality with an acceptance rate of about 80%. Still, a DBW of 1 Mbps results in good overall quality, with almost full acceptance of the participants. Excellent QoE results are attained for 8 Mbps, which shows that even if a 2 Mbps DBW allocation is high enough to reach full acceptance (cf. Figure 3), the overall experience of the user can still marginally improve. These DBW thresholds are highly important for network dimensioning, as they allow to understand the boundaries between user satisfaction and over-provisioning of resources. Very interesting is the fact that these QoE requirements in terms of DBW are more restrictive than those we found in [3] for laptops about 2 years ago, evidencing how the Facebook application has been evolving in time, becoming more network resources demanding.

B. *QoE in Mobile Web Browsing*

Figure 4 reports the overall quality and acceptability results obtained for the News website browsing tests. Note first how the quality increases in a logarithmic fashion with increasing values of the DBW. Good experience (MOS $\approx$ 4) is obtained for a DBW of 2 Mbps, and only slight QoE differences are obtained when increasing the bandwidth to up to 16 Mbps, going to MOS $\approx$ 4.15. Going in the DBW decreasing direction, the slowest tested condition still results in fair quality (MOS $\approx$ 3.5) and high acceptance rate, close to 90%.

C. QoE in Gmaps Mobile

Figure 5 reports the overall quality and acceptability results obtained for the Gmaps tests. Figure 5(a) shows that a DBW of 4 Mbps results in near optimal QoE (MOS $\approx$ 4.5), and from this value on, QoE saturation already occurs. This means that no major QoE improvements are then obtained for additional bandwidth provisioning. A DBW of 2 Mbps provides good quality results and almost full acceptance, but a DBW of 1 Mbps rapidly brings Gmaps into bad user experience.

D. QoE in WhatsApp

Figure 6 shows the QoE results for different DBW values. Users tolerate WhatsApp downloads with a good overall experience and high acceptability as long as the DBW is above 2 Mbps, but user experience heavily degrades for slower connections, resulting in very bad quality for a DBW of 500 kbps. In this case, a DBW threshold of 2 Mbps permits to approximately discriminate between good and bad experience. Given the file size used in the tests (5 MB), there is a clear saturation effect after 4 Mbps, as QoE does not increase for higher DBW values. Finally, even if the obtained results are partially biased by both the specific file size used in the tests and the participants task briefing, obtained results are similar to those we obtained in [6] for the specific case of Dropbox file sharing, suggesting that the main take aways are potentially more generic than expected when considering file downloads, either in mobile devices or in fixed ones.

V. WhatsApp Traffic and QoE in the Wild

We now tackle a different dimension of the QoE provisioning in mobile devices problem, by taking into account large-scale traffic measurements in a real cellular setting with thousands of customers. The analysis we do next considers only the case of WhatsApp, as a means to exemplify the application of some of the obtained results in the large-scale.

So far we considered the downlink throughput as the main KPI (Key Performance Indicator) reflecting QoE. However, in order to better understand the impacts of file transfer throughputs on the experience of the users downloading files with WhatsApp, the next results map downlink throughput and file size into waiting times. Download time is in fact the most relevant feature as perceived by the user when analyzing file transfers [4], as this is directly linked to anxiety and satisfaction.

According to our measurements, file downloads in WhatsApp are carried on single bulk-flow transfers, thus a single flow is observed for each single downloaded object (i.e., a videos, a photo, etc.). Therefore, the 5MB file size considered in the lab tests results in a 5MB flow in the real traffic measurements. While it is clear that the 5MB flow size reflects only a fraction of the total flows in WhatsApp (as we see next), the performed mapping permits to have some rough ideas of what the large population of WhatsApp users perceive of the service in terms of quality in the real cellular network.

A. WhatsApp Measurements Description

For the analysis of WhatsApp QoE in a real cellular network, we conducted passive measurements at the core of a European national-wide cellular network during one week in early 2014. The complete dataset consists of more than 150 million WhatsApp flows. WhatsApp uses encrypted communications, thus it is not straightforward to track its traffic in the large-scale from in-network passive measurements. We follow the approach presented in [8] to unveil and filter all the WhatsApp flows from the monitored traffic.

Flows are captured at the well-known Gn interface [11], using the METAWIN cellular network monitoring system [10]. To preserve user privacy, any user related data are anonymized, while packets' payload is removed on the fly. Traffic flows are continuously imported and analyzed through DBStream [9], a data stream warehouse tailored for large-scale traffic monitoring applications.

B. WhatsApp QoE in the Wild

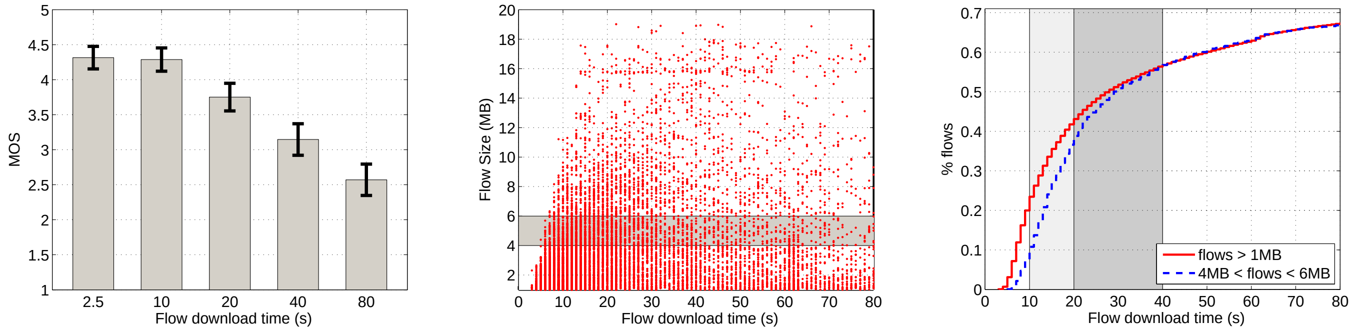
Figure 7(a) shows the QoE results for different DBW values, translated into waiting times. As an analogy to those results reported in Figure 6, the figure shows that users tolerate transfers of up to 20s long with a good overall experience, whereas transfers lasting more than 80s are considered as very bad quality. A threshold of about 40s permits to approximately discriminate between good and bad experience.

Figure 7(b) plots the Flow Size vs. the Flow Download Time (FDT) for the large-scale measurements dataset, considering only flows bigger than 1MB. If we focus on the range of flows with sizes around 5MB (flagged by gray strips in Figure 7(b)), we see that while the majority of the flows have a FDT below 40s, there are many downloads which highly exceed this threshold. Indeed, Figure 7(c) shows the distribution of the FDTs, both for all the flows with size between 4MB and 6MB, as well as for all the flows bigger than 1MB. From these CDFs, one can say that almost 40% of the WhatsApp downloads with size between 4MB and 6MB have a FDT lower than 20s, resulting in good user experience. About 60% still result in an acceptable quality, and about 35% are potentially badly or very badly perceived.

Finally, if we now assume that users are generally non experts and that file sizes are not taken into account into their quality expectations when downloading a video or a song through WhatsApp, we could say that similar results are observed for the complete dataset of downloaded flows bigger than 1MB. Of course this last observation is rather controversial, but still presents some notions on the experience of the end users.

VI. Concluding Remarks

Smartphones are becoming the Internet-access devices by default, and more and more users are accessing the most popular mobile services in their phones. In this context, network operators must understand how to manage and dimension their networks in order to correctly provision such services, avoiding



(a) WhatsApp overall quality vs. download time. (b) Flow size vs. download time. (c) Flow download time distribution.

Figure 7. WhatsApp QoE in the wild, applying lab results to a large-scale traffic measurement campaign. Figs. (b) and (c) consider only flows bigger than 1MB, to improve the stability of the download times as measured in the cellular network. Users tolerate transfers of up to 20s long with a good overall experience, whereas transfers lasting more than 80s are considered as very bad quality.

wasting additional unnecessary resources while keeping end users happy, and most importantly, reducing the chances of churning due to quality dissatisfaction. We believe that QoE has the potential to become the next guiding paradigm for managing quality provisioning and applications' design in cellular networks and mobile devices, and conducted an study shedding light in this direction.

We have presented an overview on the QoE of different services and applications with different network-level QoS requirements for the specific case of smartphone devices. The results presented in this paper are highly relevant to future 5G design and LTE evolution in better understanding the mapping between network performance and customer experience. Indeed, our results are very practical and have a paramount impact on the operation and management of mobile networks.

Finally, we have conducted a large-scale measurement campaign in an operational cellular network and mapped the lab QoE results into the obtained measurements, revealing that a non-negligible fraction of users might actually experience bad quality when using WhatsApp as a means to share files with other friends, specially in those cases where the contents are to be accessed fast, for example when sharing some top-popular content between friends. As future direction, we shall extend our study by adding application-layer measurements into the analysis; we are currently developing new monitoring tools to capture metrics such as page load times for the case of web browsing and Facebook.

Acknowledgments

The research leading to these results has been done within the project ACE 3.0 at the Telecommunications Research Center Vienna (FTW), and has received partial funding from the European Union under the FP7 Grant Agreement n. 318627, "mPlane".

References

[1] Cisco Visual Networking Index, "Global Mobile Data Traffic Forecast Update 2014-2019 White Paper", Dec. 2014.

[2] eMarketer Newsletter, "2 Billion Consumers Worldwide to Get Smart(phones) by 2016", Dec. 2014.

[3] P. Casas et al., "YouTube & Facebook Quality of Experience in Mobile Broadband Networks", in *IEEE Globecom Workshops*, 2012.

[4] P. Casas et al., "Quality of Experience in Cloud Services: Survey and Measurements", in *Computer Networks*, vol. 68, pp. 149-165, 2014.

[5] M. Jarschel et al., "An Evaluation of QoE in Cloud Gaming Based on Subjective Tests", in *FINGNet*, 2011.

[6] P. Casas at al., "A First Look at Quality of Experience in Personal Cloud Storage Services", in *IEEE ICC Workshops*, 2013.

[7] P. Casas et al., "Quality of Experience in Remote Virtual Desktop Services", in *IFIP/IEEE QCMan Workshop*, 2013.

[8] P. Fiadino et al., "Vivisecting WhatsApp through Large-Scale Measurements in Mobile Networks", extended abstract in *ACM SIGCOMM*, 2014.

[9] A. Bär, P. Casas, L. Golab, A. Finamore, "DBStream: an Online Aggregation, Filtering and Processing System for Network Traffic Monitoring", in *TRAC*, 2014.

[10] F. Ricciato, "Traffic Monitoring and Analysis for the Optimization of a 3G network", in *IEEE Wireless Communications*, vol. 13(6), 2006.

[11] J. Bannister, P. Mather, S. Coope, "Convergence Technologies for 3G Networks: IP, UMTS, EGPRS and ATM". Wiley, 2004.

[12] P. Fiadino et al., "Vivisecting WhatsApp in Cellular Networks: Servers, Flows, and Quality of Experience", in *TMA*, 2015.

[13] I. Ketykó et al., "QoE Measurement of Mobile YouTube Video Streaming", in *MoViD*, 2010.

[14] G. Gómez et al., "YouTube QoE Evaluation Tool for Android Wireless Terminals," in *EURASIP Journal on Wireless Communications and Networking*, vol. 164, 2014.

[15] V. Aggarwal et al., "Prometheus: Toward Quality-of-Experience Estimation for Mobile Apps from Passive Network Measurements," in *ACM HotMobile*, 2014.

[16] Int. Telecommunication Union, "Methods for Subjective Determination of Transmission Quality", *ITU-T Rec. P.800*, 1996.

[17] Int. Telecommunication Union, "Subjective Video Quality Assessment Methods for Multimedia Applications", *ITU-T Rec. P.910*, 2008.

[18] Int. Telecommunication Union, "Subjective Testing Methodology for Web Browsing", *ITU-T Rec. P.1501*, 2013.

[19] Int. Telecommunication Union, "Estimating End-to-End Performance in IP Networks for Data Applications", *ITU-T Rec. G.1030*, 2014.

[20] S. Hemminger, "Network Emulation with NetEm", in *Proc. LCA*, 2005.