Towards Evaluating Type of Service Related Quality-of-Experience on Mobile Networks

Christos Tsiaras*, Anuj Sehgal†, Sebastian Seeber‡, Daniel Dönni*, Burkhard Stiller*, Jürgen Schönwälder† and Gabi Dreo Rodosek‡
*University of Zürich, Communication Systems Group, Binzmühlestrasse 14, 8050 Zürich, Switzerland
Email: {tsiaras, doenni, stiller}@ifi.uzh.ch
†Computer Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany
Email: {s.anuj, j.schoenwaelder}@jacobs-university.de
‡Universität der Bundeswehr, Faculty of Computer Science, 85577 Neubiberg, Germany
Email: {sebastian.seeber, gabi.dreo}@unibw.de

***Abstract*—Quality of Service (QoS) metrics have been traditionally used to evaluate the perceived quality of services delivered by network operators. However, these metrics are not suitable for evaluating the experience of an end-user. The experience of a user is quantified based upon activities such as speed of web page loading, quality of video streaming, or voice quality of Internet-telephony. Due to the temporal and geographical nature of mobile networks, the perceived experience of a user may change based on location and time. Mobile operators may prioritize certain services over others, leading to a service type dependent Quality of Experience (QoE). In this paper we present a mobile application developed to gather metrics necessary to evaluate QoE in a mobile environment. Our approach towards obtain not just a general, but service specific mean opinion score (MOS) to quantify QoE is also discussed. Initial experiments and measurement tests show that it is possible for the same operator to deliver different QoE based on service and even time-of-day.**

Keywords—Quality of Experience (QoE), Mean Opinion Score, Mobile Network Operators, Traffic Shaping, Mobile Networks, Voice over IP (VoIP), Android, Internet Measurement

I. INTRODUCTION

Emergence of high-speed mobile networks based on Universal Mobile Telecommunications System (UMTS), Long Term Evolution (LTE) and other similar standards is leading to increased usage of Internet based services. Availability of higher throughput allows users to go beyond simple web-browsing on mobile phones by enabling services such as video streaming and Voice over IP (VoIP). However, given the high costs of acquiring spectrum necessary for delivering such services it is possible that Mobile Network Operators (MNOs) limit the actual data-rate achievable by users [1]. At the same time, increasing competition and affordable services are growing subscriber numbers, thereby putting additional strain on mobile networks [2]. Increased network load can lead to dynamically changing local congestion, which causes decreased channel availability, thereby causing degradation of service quality or compromised sessions [3]. Both scenarios, throttled data-rate and local congestion, can severely impact user experience.

It is also important to keep in mind that the expectations of mobile network users are based upon their experiences with fixed-line networks, which are traditionally more stable and less congested [1]. To offer good user experiences a mobile network needs to be well planned and optimized, which requires the MNO to understand the traffic characteristics, especially due to the geographical and dynamic nature of traffic in mobile networks. Furthermore, to appropriately engineer the network, it is important for MNOs to understand the geographical and temporal, type of service related, Quality-of-Experience (QoE) from the users' perspective.

A service related QoE can be quantified based upon the Mean Opinion Score (MOS) value, shown in Table I, that represents the subjective experience of a user for a particular quality of the network. There already exist a number of MOS tests for evaluating video streaming [4], telephony [5] and VoIP quality [6], [7]. However, these do not quantify the overall QoE using multiple possible metrics, given time and location, within a mobile network. Furthermore, while these studies present approaches for estimating the MOS, none of them provide a method for gathering metrics from within a mobile network. Being able to collect data from a users' actual device can also allow derivation of a generic MOS value that takes into account different service based QoE.

To determine the user perceived service related QoE, quantified by MOS, we developed an Android application that records various network conditions at the time a user executes a test. The results from the Android application, along with a few responses from users, are used to obtain a location, time and service type based MOS. This MOS will in the future be displayed on a global map that relates MOS values to colors for easy reading. Usage of such a map can assist users in comparing the QoE provided by different MNOs; operators and regulators could use it to discover specific geographical areas that might need network optimization to improve service of a particular type.

The remainder of this paper is structured as follows. Related work is discussed in Section II, followed by the architecture of the Android application in Section III. The overall MOS calculation model is presented in Section IV and Section V presents a discussion on how appropriate parameters for obtaining the MOS are chosen. An experimental setup and results gathered from some measurement tests of

978-1-4799-3060-9/14/$31.00 © 2014 IEEE

Table I. THE MOS SCHEME RECOMMENDED BY THE ITU-T [14]

MOS Value	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

various services in networks of multiple MNOs are provided in Section VI. Finally, Section VII draws conclusions and presents future work.

II. RELATED WORK

Multiple studies investigate MOS for specified applications like VoIP [6], [7], traditional telephony [5] or video streaming quality [8]. These studies do not quantify QoE related to time and location within a mobile network, which can be useful in studying the evolution of a network or even load characteristics.

In mobile environments there exist a few measurement platforms which take into account a single value showing these on coverage maps to allow a location based comparison of providers. The projects OpenSignal [9] and Epitiro [10] focus solely on collecting and representing signal strength. On the other hand, RootMetrics [11] and Netradar [12] go further and also collect throughput information. Cisco GIST [13] shows throughput coverage, however is limited to the USA. The problem with many of these tools is that either they are proprietary or the raw data is limited and not available for examination.

Some work has also been invested in building measurement platforms that record location, time [15], [16], speed [15], signal strength [17] or used radio technology [18], [19] information. Others investigate the affects of hand-over between cellular towers [20], [21]. However, all of them are limited to a study of general throughput, and not service or protocol specific throughput and latency in mobile networks. This work here focuses on obtaining MOS values for popular activities on smartphones by gathering, via an Android application, and analyzing the appropriate metrics related to the protocols used to deliver these services.

III. MEASUREMENT APPLICATION

Measuring voice quality on phone calls placed in a mobile network is no longer enough since smartphones are capable of performing data-intensive tasks. Users may choose to watch videos online, share pictures, use instant messaging or VoIP to reduce expenditure. Increase in demand for such services can overload existing mobile network access infrastructure, which leads to degraded user experience. Furthermore, in light of a channel plagued with low bandwidth and reliability issues, compounded by high saturation in some areas, MNOs might deploy traffic shaping policies in order to improve the average customer experience. However, this average experience, while better than the worst-case, might still be below the QoE expected by customers.

It is, as such, important to take QoE measurements from smartphone users in order to understand the mobile network experience from their perspective. While there exist many throughput measurement applications for smartphones, none perform protocol based measurements or measure QoE [22], as they are limited to collecting basic network information and performance data. Traditional desktop applications are not useful because they do not function on smartphones.

Table II. ANDROID APPLICATION QOE TEST OPTIONS

Choice	Protocols Tested
Web Browsing	HTTP
Internet Telephony (VoIP)	SIP, RTP
Streaming Media	Flash Video, RTSP

This work extends BonaFide [22], an open source Android application designed to detect traffic shaping in mobile networks. BonaFide provides a good basis for deriving QoE representation in MOS values because the application is already designed to measure performance of multiple protocols. Furthermore, support for additional protocols can be added via protocol description files.

A. System Architecture

BonaFide uses a client-server architecture to perform measurement tests. The client, *i.e.*, the Android application, gathers network performance metrics by connecting to and sending data to the server. Since it is possible for MNOs to have protocol based traffic shaping policies applied, just measuring throughput of a link using random data is not enough. As such, BonaFide presents users a list of protocols they would like to test. The protocol description file for the chosen protocol is then downloaded from the server. These files contain a set of rules that define the client and server behavior [22]. A protocol description file for HTTP looks as below:

```
protocol HTTP
PFport 30008
RFport 31008

request string("GET /wiki/Computer_Network HTTP/ 1.1")
byte(13) byte(10) string("Host: en.wikipedia.org") byte(13)
byte(10) string("User-Agent: Mozilla/5.0 (Linux; U; Android
2.3.5; en-de; HTC Desire S Build/GRJ90) AppleWebKit/ 533.1
(KHTML, like Gecko) Version/4.0 Mobile Safari/533.1")
byte(13) byte(10) string("Accept: text/html") byte(13)
byte(10) string("Connection: close")
byte(13) byte(10) byte(13) byte(10)

response string("HTTP/1.1 200 OK") byte(13) byte(10)
string("Server: Apache") byte(13) byte(10) string("Content
-Language: en") byte(13) byte(10) string("Content-type:
text/html; charset=utf-8") byte(13) byte(10) string("Content
-Length: 20") byte(13) byte(10) byte(13) byte(10)
string("12345678901234567890")
```

As can be seen above, the protocol description files contain the protocol name, port numbers to be used for tests and any well known ports associated with the protocol. Instructions for the client application regarding construction of payloads conforming to the chosen protocol follow. In the provided example, the client is instructed to create traffic that contains a `GET` request for a web-page from Wikipedia. The response

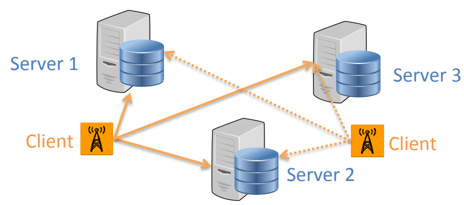

Figure 1. The modified BonaFide system architecture.

from the server will contain an `HTTP 200 OK` along with a 20 *byte* payload.

To avoid backbone networks influencing the results, we endeavor to place measurement servers as close to the client as possible. This is achieved by having multiple servers geographically distributed (c.f. Figure 1). During each measurement test, some basic network information is recorded by the application. BonaFide only gathers protocol and random data related throughput, and so it was extended to also gather metrics such as network operator and SIM provider ID to determine whether the network is a Mobile Virtual Network Operator (MVNO) or not; signal strength to determine whether results are dependable; latency to ensure that the network is actually usable in scenarios when high throughput might be possible but latency makes VoIP calls impossible or web-page load times too high; and the current location of the measurement device to aid the development of a location based QoE database.

Depending on the chosen protocol, tests are either bi-directional (*e.g.*, VoIP), *i.e.*, upload and download performance of the link, or unidirectional (*e.g.*, Flash Video). Normally BonaFide performs only bi-directional tests, but this modification was made in keeping with the general usage scenarios of these protocols and to save upon costly mobile data plans. Collected results are stored in a database on the server that the client initiated a connection to, from where they are replicated across all measurement servers for later retrieval.

B. Measurement Test Lifecycle

At the outset, the application determines the smartphone's current location to determine the closest server it should connect to, from where the list of supported protocols/services is downloaded. The user is then presented with a choice of tests to run, which can be seen in Table II. These simplified choices are presented instead of protocol names, unlike in the original BonaFide application, since users can be non-technical. Users may choose a single service category, or multiple ones for the measurement.

It then proceeds to record the aforementioned network performance metrics. Following this, bi-directional or unidirectional protocol based throughput calculation is performed by using the protocol definition files to generate traffic that mimics the behavior of chosen protocols. Results are uploaded to the server, for offline processing to generate representative MOSs. Figure 2 provides an overview of the entire measurement test lifecycle. As a verification step, the application also asks the user to provide feedback concerning their experience with the MNO. Optional information regarding the service plan is also requested from the user.

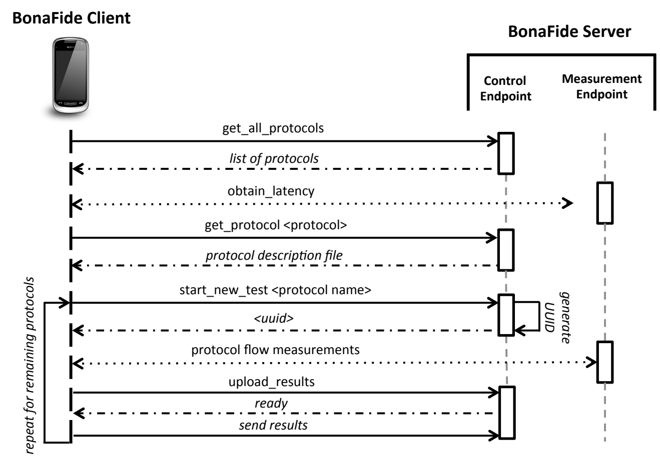

Figure 2. Lifecycle of the modified BonaFide measurement test.

IV. Mean Opinion Score (MOS) Calculation

The MOS is a number that reflects the end-users' QoE. The Telecommunications Standardization Sector of the International Telecommunications Union (ITU-T) has defined in recommendations P.800 [14], P.800.1 [23] and P.805 [24], a five-point scale that illustrates the QoE of the end-user. The ITU-T MOS scale is summarized in Table I and it is used here for all MOS calculations.

Furthermore, this paper introduces a generic MOS calculation method, which encapsulates the affect of diverse parameters, in services, such as video streaming, VoIP, browsing etc. Previous work [12] presents the raw-data concerning measurements in selected parameters, such as throughput, delay, jitter etc. However, those works do not deliver an overall picture when it comes to the expected quality of a specific service. Focusing in one parameter might generate a misleading expectation since enough throughput do not necessarily mean that the QoE of the end-user during a VoIP session will be high, because maybe at the same time high jitter will result a reduced QoE. Thus, it is essential to consider multiple parameters when calculating the MOS for a specific service as a total.

A. Specific MOS

Some increasing parameters, such as the available throughput, affect the QoE positively. In this work those parameters are referred to as *increasing* parameters. On the other hand, parameters such as the delay, or jitter affect negatively the QoE while increasing. Those parameters termed in this work *decreasing* parameters. Nevertheless, for every parameter there is an *expected* value that either the end-user is paying for, or the Service-level Agreement (SLA) defines, or a service demands to perform as expected. *E.g.*, Hulu recommends a downstream throughput of at least 1.5 $Mbps$ [25] for smooth playback experience of Standard Definition (SD) 480p videos.

Let x be the value of an *increasing* parameter i and x_0 the *expected* value of the same parameter. Let $e_i(x)$ be the MOS that reflects the QoE considering the parameter i according to Table I. Finally, let $e_i(x_0) = 4$ since the *expected* value reflects

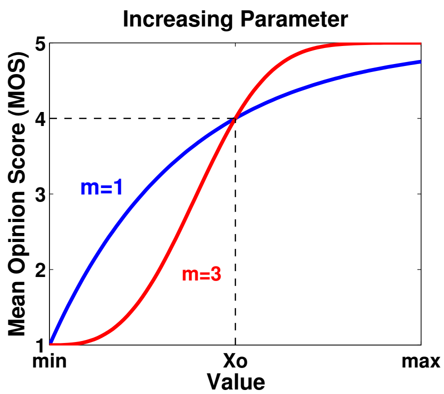

Figure 3. MOS for *increasing* Parameters for Two Different Influence Factors

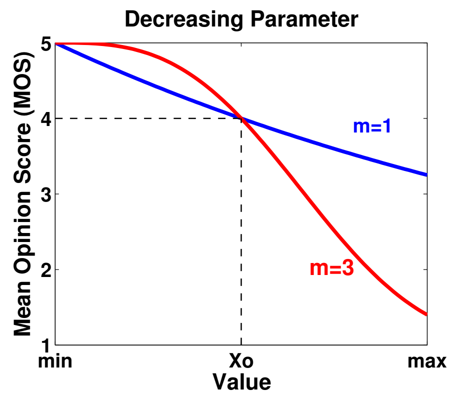

Figure 4. MOS for *decreasing* Parameters for Two Different Influence Factors

an overall good experience of a service. For the *increasing* parameters, the MOS $e_i(x)$ tends asymptotically to 5 when x is increasing since this is the upper limit of the MOS scale. On the other hand, when the value x is minimum, the MOS $e_i(x)$ is equal to the lower bound of the MOS scale which is 1. Thus, Equation 1 is used to calculate the MOS of an *increasing* parameter i. Equation 2 is used to calculate the MOS of a *decreasing* parameter j. In the case of *decreasing* parameters the MOS $e_j(x)$ is 5 when the value x is minimum and when the value x is increasing the MOS $e_j(x)$ tends asymptotically to 1. Similar with the *increasing* parameter case, the MOS of the *expected* value $e_j(x_0) = 4$.

Last but not least, let $m \in \mathbb{R}_0^+$ be the influence factor of a parameter. The influence factor represents how fast the fluctuation of the value x affects the MOS. Figure 3 and Figure 4 illustrate the MOS of a) an *increasing* and b) a *decreasing* parameter respectively, for two different influence factors $m = 1$ and $m = 3$. It is shown that the higher the influence factor m, the larger the MOS variation becomes for a given fluctuation of the value x.

$$
\begin{aligned}
e_i(x) &= 4(1 - e^{-\lambda x^m}) + 1 \\
\lambda &= x_0^{-m} \ln 4
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
e_j(x) &= 4e^{-\lambda x^m} + 1 \\
\lambda &= x_0^{-m} \ln \frac{4}{3}
\end{aligned} \tag{2}
$$

B. Generic MOS

Each service has unique characteristics and demands to perform satisfactorily. Some services, such as video streaming, are more tolerant to jitter than real-time services, such as video conferences, or VoIP calls. However, both services require sufficient throughput. Thus, there is no universal way that the fluctuation of each parameter affects the overall QoE of a service.

Combining the MOSs of a set $X = \{x_1, ...x_k, ...x_N\}$ of $N \in \mathbb{N}^+$ diverse parameter values x_k to one generic MOS $E(X)$ demands to consider a different weight w_k for each parameter k, since the importance of each parameter might be different. Equation 3 calculates the total MOS for the set X respecting the ITU-T scale. The weights $w_k \in [0, 1]\ \forall\ k$ represent the contribution, in terms of percentage, of each parameter k in the generic MOS $E(X)$. Thus, $\sum_{k=1}^{N} w_k = 1$ since the total summary of all the contribution weights cannot be different than 100%. Figure 5 illustrates the generic MOS of two parameters, one *increasing* and one *decreasing* with influence factors $m = 1$ and $m = 3$ respectively. In this example the contribution weight of both parameters is selected to be 50%. The white area on the graph marks the pairs of the parameters values that result in a MOS of 4.

$$
E(X) = 1 + \prod_{k=1}^{N} [e_k(x_k) - 1]^{w_k} \tag{3}
$$

To generate the generic MOS in each type of service the following information has to be defined for each parameter k that affect the total score. First of all, each parameter need to be identified as *increasing*, or *decreasing*. Then, the minimum and maximum and the *expected* value x_0, as well as the influence factor m for each parameter has to be defined. The influence factor m can also be different for each parameter's values $x \leq x_0$ and $x \geq x_0$ since the positive and the negative influence intensity may be different. Last but not least, the weight w_k of each parameter k in the generic MOS need to be defined.

V. Determining MOS Calculation Parameters

It is important to identify the network characteristics and parameters that have a pronounced affect upon the QoE for general data traffic and a selected list of popular services within mobile networks, as seen in Table II. These parameters, such as uplink or downlink throughput and latency, can be characterized as either *increasing* or *decreasing*, based on whether a larger value leads to a positive or negative impact on the QoE.

Defining the maximum and minimum performance possible of a protocol, within the realm of perceived impact on QoE by the user, is important to calculate the appropriate MOS. The parameter x_0, which indicates the expected performance to achieve a MOS of 4 (good), also needs to be defined for each protocol, along with that influence factors m^- and m^+, which determine the rate of decrease or increase of MOS below and above x_0. Finally, the weight w_k of each service, as it

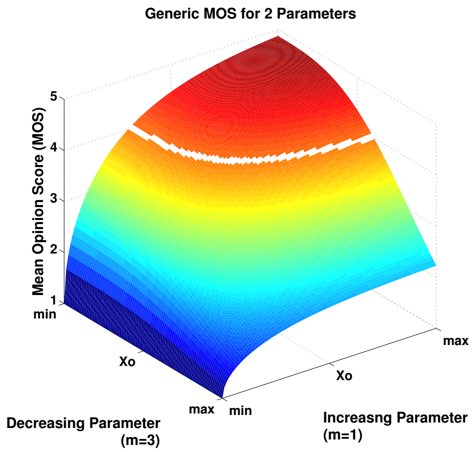

Figure 5. Generic MOS for one *increasing* and one *decreasing* Parameter with Different Influence Factors (m=1 and m=3 respectively)

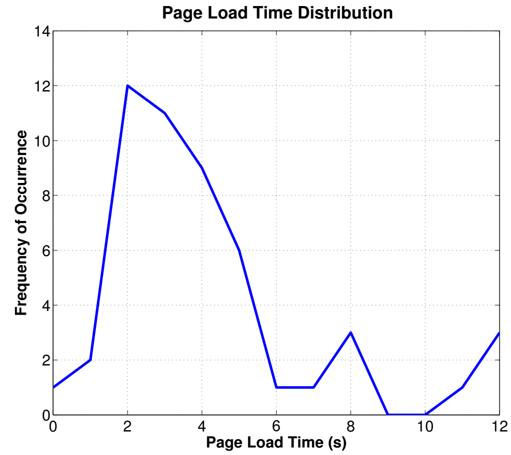

Figure 6. Distribution of the mobile webpage load times for the top 50 web-sites, *i.e.*, the number of web-sites that load within a specific time.

contributes to obtaining a generic overall MOS needs to be selected taking into account user perspective on the importance of each service category on their QoE.

A discussion regarding how values for each of these parameters was chosen for the considered service categories is provided in this section, with Table III summarizing the values chosen for each parameter relative to the service category.

A. General Considerations

Throughput and latency are parameters that generally have a significant affect upon QoE, and by extension the MOS. In some cases, *e.g.*, VoIP telephony, bi-directional throughput affects the overall MOS, since data must travel in both directions. However, in other cases, *e.g.*, video streaming and web-browsing, the downlink throughput is more important. In all cases, the minimum value of the available throughput which corresponds to the lowest MOS $e(throughput) = 1$, irrespective of link direction is $0\,Mbps$, which represents no connectivity.

The maximum throughput attainable is governed by the mobile link technology. The average sector throughput in LTE MIMO 4x4 with 20 MHz bandwidth, the most deployed form of LTE, provides a maximum of 12.7 Mbps uplink and 50.1 Mbps downlink throughput [26]. These maximum and minimum values are the same for all service types due to their direct dependence on access technologies.

The influence factors m (m^- for $x < x_0$ and m^+ for $x > x_0$) are calculated for all service categories based on the rate of increment of the MOS from 4 to 5 and the decrement from 4 to 3, given the rate of change of the *increasing* or *decreasing* parameter. For example, a 25% reduction of x_0 for an increasing parameter i (cf. Equation 1) should yield $e_i(0.75 \cdot x_0) = 3$, as such $m^- = 2.41$. For the same parameter if a 100% increase of x_0 should result in $e_i(2 \cdot x_0) = 3$, then $m^+ = 2.58$. Similarly, for a decreasing parameter j (cf. Equation 2), if a 50% reduction of x_0 results in $e_j(0.5 \cdot x_0) = 5$, then $m^- = 10.17$ and if a 15% increase of x_0 results in $e_j(1.15 \cdot x_0) = 3$, then $m^+ = 6.29$.

B. Web Browsing

Mobile web-browsing experience is closely related to the achievable HTTP downlink throughput and the size of webpages being browsed. To establish the necessary value for x_0, the size of web pages for the 50 most popular web-sites [27] was obtained. These results indicated that the average size of webpages for the top 50 visited web-sites is $1881kB$[1]. Many web-sites also provide webpages optimized for reduced data consumption on mobile devices, taking these into account as well reduced the average size of webpages to $720\,kB$[1].

A test to determine the load times for home pages of these top 50 visited web-sites was carried out on mobile networks; mobile phone optimized webpages were loaded where possible. The test was carried out on UMTS networks since most mobile subscribers have access to them. To obtain a meaningful result, the load times were rounded up or down to their closest integer representation. Following this, the frequency of occurrence for each load time was obtained, and can be seen in Figure 6.

The weighted mean of the rounded load times was obtained, by using the respective frequency of occurrence as weight. This determined that it takes approximately $4.33\ s$ to load average sized webpages on UMTS mobile networks. Since the top 50 web-sites are likely to account for bulk of the mobile Internet traffic, this means users are likely to be happy with the performance of these webpages, thereby making approximately $4.33s$ the expected load time for a MOS score of 4. This translates to a throughput of $x_0 = 1330\,Kbps$, based on an average webpage size of $720\,kB$.

Latency can also have an impact upon the time it takes to load a webpage, and therefore on the QoE. A method similar to that used for determining the throughput x_0 was used to obtain the latency x_0 as well. The weighted mean for latency rounded to the nearest $50\ ms$ was determined to be $x_0 = 523\,ms$. The minimum latency is set to $1\ ms$ and maximum to $2000\ ms$, since it is not expected to lie outside this range and the connection still be considered

[1] As per the HTTP Archive [28] dataset from 2014-01-15.

Table III. MOS-RELATED VALUES

Service	Protocol(s)	Effect	Parameter	min	max	x_0	$MOS = 3$	$MOS = 5$	m^-	m^+	w_k
Browsing	HTTP	increasing	downlink throughput	$0\,Mbps$	$50.1\,Mbps$	$1330\,Kbps$	-25%	$+100\%$	2.41	2.58	75%
		decreasing	latency	$1\,ms$	$2000\,ms$	$523\,ms$	$+15\%$	-50%	10.17	6.29	25%
Video	Flash, RTSP	increasing	downlink throughput	$0\,Mbps$	$50.1\,Mbps$	$1.5\,Mbps$ 480p	-20%	$5\,Mbps$ 720p	3.11	1.49	100%
VoIP	SIP, RTP	increasing	uplink throughput	$0\,Mbps$	$12.7\,Mbps$	$8\,Kbps$[3]	$5.3\,Kbps$[2]	$64\,Kbps$[3]	1.67	0.86	25%
		increasing	downlink throughput	$0\,Mbps$	$50.1\,Mbps$	$8\,Kbps$[3]	$5.3\,Kbps$[2]	$64\,Kbps$[3]	1.67	0.86	25%
		decreasing	latency	$1\,ms$	$2000\,ms$	$150\,ms$[4]	$+50\%$[3]	-50%	10.17	2.16	50%

usable. Also, increasing throughput and decreasing latency will improve MOS. Throughput is chosen as being three times more important, *i.e.* w_k, than latency in the overall MOS calculation because the effects of throughput are always felt, whereas only a really poor latency will lead to a significant impact on page load times.

C. Video Streaming

Video streaming is a common task performed on smartphones, and so to obtain MOS values, the downlink throughput of RTSP and Flash Video protocols is considered. Uplink throughput is not important since video streaming is a download intensive service. Most modern smartphones are equipped with screens capable of displaying 720p or better quality high-definition video. As such, users expect at least 480p standard definition quality video to categorize it as good quality [8], [29]. The throughput requirements for popular video qualities can be seen in Table IV. Based on these values and the fact that 480p video quality corresponds to a MOS of 4, $x_0 = 1024\,Kbps$. Increasing throughput slowly improves QoE because significantly higher throughput is necessary to sustain high definition video streaming, while reducing throughput quickly decreases the MOS due to lack of bandwidth to sustain acceptable quality. Latency does not have a direct impact upon video quality, even though a high latency can lead to choppy video playback. But this is easily solved with a larger buffer in the video playback application, and as such only throughput has a $w_k = 100\%$.

D. VoIP Telephony

To quantify VoIP QoE in mobile networks, performance of SIP and RTP traffic is evaluated. Due to the numerous studies on MOS for VoIP, it is already known that the G.729 codec, which needs $8\,Kbps$ of bi-directional throughput, corresponds to MOS of nearly 4 [30], thereby making $x_0 = 8\ Kbps$. The G.711 codec delivers the best MOS at a throughput requirement of $64\ Kbps$. Since phone calls carry audio data in both directions, improving throughput leads to better QoE, whereas reducing bandwidth causes the MOS to drop quickly.

Another factor that affects quality of VoIP calls is latency, a high value of which can lead to the existence of a lag and a low QoE. The minimum and maximum latency are once again set to $1\,ms$ and $2000\,ms$. This maximum latency is four times larger than the round-trip-time experienced while using a geostationary satellite for communication [31]. If a packet is delayed for longer than this period it could be considered lost as well. Previous studies have shown that a maximum one-way delay of $150\,ms$ is tolerable to humans during VoIP conversations [32], thereby leading to a $x_0 = 150\,ms$. Only a decreasing latency can improve QoE, while a large one has a significantly negative impact on the MOS value.

VoIP conversations are quite sensitive to fluctuations in throughput and latency, as such their contribution in obtaining an overall VoIP MOS value is equal. Also, the throughput needs to provide similar performance in both, uplink and downlink, directions for acceptable call qualities. This leads to a w_k of 25% for each direction of throughput and 50% for latency.

VI. EXPERIMENTAL RESULTS

Having chosen appropriate values for x_0, m^-, m^+ and w_k for the service categories seen in Table III, we ran a few measurement tests across different operators to obtain an initial evaluation of QoE on UMTS networks. For the purposes of this evaluation measurement servers were setup in three geographical locations, *i.e.*, Bremen, Germany; Munich, Germany; and Zürich, Switzerland. Protocol description files for HTTP, Flash Video, RTSP, SIP and RTP were installed on the server.

Networks of three MVNOs from Germany and one MNO each in Belarus and the USA were used to perform the measurement tests. The tests on German MVNOs were performed in the area around the campus of Jacobs University Bremen, while the tests in Minsk, Belarus and San Francisco, USA were carried out within the downtown region. During this initial evaluation phase, the user was requested to choose all service categories to run the measurement tests on. Each measurement was repeated five times during a single run of the test in order to obtain higher confidence in results. The experiment was carried out at 9 AM and 7 PM, in order to test if time of day would make a difference in the service category based QoE. This meant that for each MNO/MVNO there were a total of 10 measurements per service category obtained. Initially, the tests were only performed on UMTS networks. For the tests within Germany, the BonaFide server located in Bremen, Germany was used as the measurement endpoint since it was closest to the client locations. The same server was evaluated as being geographically closest for the tests performed from Belarus as well. This server was also used for measurements from the USA due to its proximity to the DE-CIX and AMS-CIX, two of the largest Internet exchanges on mainland Europe.

The general service category MOS values were obtained for each of the 10 measurements. Average MOS values rounded

[2] G.723.1 is the codec that demands the lowest possible throughput for a VoIP call [33] ($5.3\,Kbps$ using a $20\,byte$ frame).

[3] The maximum throughput demands for a codec nowadays is $64\,Kbps$ [30]

[4] ITU-T, One-way Transmission Time. ITU-T Recommendation G.114 [32]

Table IV. MINIMUM REQUIRED THROUGHPUT TO MAINTAIN A PARTICULAR VIDEO QUALITY [25], [34]

Video Quality	Throughput [kbps]
SD 360p	1024
SD 480p	1536
HD 720p	5120
HD 1080p	8192

to single decimal point precision were chosen as the MOS representative of that service category since the standard deviation was never found to be above 0.02 for any category. To establish a comparison, the same measurements were also repeated on EDGE networks of the MNO/MNOs. The per service category MOS results for UMTS and EDGO networks of the tested operators can be seen in Table V.

The UMTS networks of almost all operators seem to provide services that can be classified as almost good or above, except for AT&T. It is interesting to note that due to the different performance requirements of each type of service within the same network it is possible for VoIP QoE to be better than video streaming, which in turn might be better than web browsing. The performance of SIP and RTP on UMTS networks is excellent on all UMTS networks, except AT&T, because of the low throughput requirements of the codecs. A high-enough throughput also means that all providers, except AT&T, are able to deliver at least 480p quality standard definition streaming video. The QoE loading while loading webpages may be lower than that of QoE or video streaming because of the need to load a relatively large file very quickly, without the benefit of being able to buffer it unlike with videos. The QoE anomaly of AT&T cannot be explained sufficiently, however, it is hypothesized that it might be due to congestion in the local cell. Further testing will need to be done to establish if this is the case or not.

While the morning and evening QoE was stable across all operators, never yielding a standard deviation above 0.02, the Congstar network provided an interestingly unique case where the UMTS QoE during the morning was significantly better than during the evening. This suggests that either the cell being tested was overloaded, or the operator is applying time based traffic shaping policies. However, since the QoE of VoIP is the lowest in the evening the likelihood of traffic shaping being the reason for such behavior is higher, especially when more throughput intensive tasks like web browsing and video streaming delivered a higher QoE than VoIP. Looking at the throughput values collected by BonaFide confirmed this suspicion, since the throughput of SIP and RTP was 95% lesser than other protocols, even though the overall performance in the evening suffered compared to the morning.

The results from EDGE networks are also quite interesting. It is quite clear that the throughput provided by EDGE is not enough for tasks like video streaming and browsing since the low throughput leads to a very low QoE. However, since the throughput requirements for a high VoIP related MOS score are not much, even on a slow network, which is quite frustrating to browse webpages on, VoIP phone calls generally provide a high QoE. It is interesting to note that once again Congstar's network applies time based traffic shaping policies since a MOS of 4 for VoIP reduces to just 1 in the evening. A continued poor performance of the AT&T network,

Table V. THE SERVICE BASED MOS SCORES OBTAINED FROM TESTS CONDUCTED ON UMTS NETWORKS OF MULTIPLE OPERATORS

MNO/MVNO	Test Location	VoIP	Video	Browsing	Type
Aldi Talk	Bremen, DEU	5.0	4.0	3.9	UMTS
		5.0	1.0	1.0	EDGE
NettoKOM	Bremen, DEU	5.0	4.0	3.7	UMTS
		4.0	1.0	1.0	EDGE
Congstar *(9am)*	Bremen, DEU	5.0	5.0	5.0	UMTS
		4.0	1.0	1.0	EDGE
(7pm)		1.0	2.0	2.0	UMTS
		1.0	1.0	1.0	EDGE
MTS	Minsk, BLR	5.0	5.0	5.0	UMTS
		5.0	1.0	1.0	EDGE
AT&T	San Francisco, CA	1.0	2.0	1.0	UMTS
		1.0	1.0	1.0	EDGE

even in regards to SIP, warrants further study and gathering measurements from multiple locations in their network.

These results clearly highlight that QoE is closely tied to the type of service being used on a mobile network, and just recording the throughput of random data while using mobile applications to record crowdsourced QoE is definitely not enough. Varying traffic policies, load and other factors can easily cause divergent QoE across different service categories.

VII. CONCLUSION AND FUTURE WORK

This work presented a prototype application that evaluates the user perceived QoE in various Internet services over mobile networks. This open source Android application, BonaFide, was originally designed to study traffic shaping in mobile networks, but has been extended here to obtain metrics necessary to obtain service specific MOS values. The application has also been extended to support measurements across multiple servers, to avoid the affects of backbone networks becoming pronounced. A QoE evaluation model, that encapsulates the affect of multiple parameters on QoE of a particular service at a given time and location, following the ITU-T recommended MOS scale was also proposed. The obtained MOS reflects the QoE a user can expect for a specific type of service, while being connected to a specific MNO.

Since measuring the performance of a random data flow cannot reveal service specific QoE in a mobile network, measurement experiments have been used to gather metrics by using the extended BonaFide application, from multiple MNOs. The obtained results were mapped to service specific MOS values for the user, since presenting raw-data such as throughput, delay, signal strength etc. is meaningless. During the experimental evaluation it was discovered that at a single given location and time, it is possible to have divergent QoE for different types of services while connected to a single MNO. It was also found that time of day can have significant impact upon the QoE due to changing network loads or even time-based traffic shaping policies of a MNO. This makes it clear that it is important to not only measure a generic QoE in mobile networks, but also evaluate it based on different services.

Future calculations of MOS will also take into account other parameters such as jitter, packet-loss and signal strength. The BonaFide client will be extended to support the metrics it does not collect and capabilities to evaluate the MOS of additional traffic types, such as mobile games, will also be added. The MOS calculation parameters will be defined for the new service types as well. The user collected feedback regarding their perceived MOS will be used to fine tune the MOS calculation parameters so that an accurate representation of user perceived MOS is made available. The collected MOS values will also be plotted on a geographical map to allow easy comparison of service specific QoE across MNOs. The BonaFide application will also be extended to other popular mobile operating systems, so as to collect additional data and also improve the accuracy of QoE representations.

ACKNOWLEDGMENT

This work was partly funded by Flamingo, a Network of Excellence project (ICT-318488) and SmartenIT (ICT-317846) supported by the European Commission under its Seventh Framework Programme.

REFERENCES

[1] P. Svoboda, "Measurement and Modelling of Internet Traffic over 2.5 and 3G Cellular Core Networks," Ph.D. dissertation, Vienna University of Technology, November 2008.

[2] A. Riikonen, "Mobile Internet Usage - Network Traffic Measurements," Master's thesis, Helsinki University of Technology, September 2009.

[3] R. Goleva, D. Atamian, S. Mirtchev, D. Dimitrova, and L. Grigorova, "Traffic Sources Measurement and Analysis in UMTS," in *Proceedings of the 1st ACM Workshop on High Performance Mobile Opportunistic Systems*. ACM, 2012, pp. 29–32.

[4] M. Ries, O. Nemethova, and M. Rupp, "Reference-Free Video Quality Metric for Mobile Streaming Applications," in *Proc. of 8th Int. Symp. on DSP and Communication Systems & 4th Workshop on the Internet, Telecommunications and Signal Processing*, 2005, pp. 98–103.

[5] ITU-T, *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*. ITU-T Recommendation P.862, February 2001.

[6] M. Manousos, S. Apostolacos, I. Grammatikakis, D. Mexis, D. Kagklis, and E. Sykas, "Voice-Quality Monitoring and Control for VoIP," *IEEE Internet Computing*, vol. 9, no. 4, pp. 35–42, 2005.

[7] R. G. Cole and J. H. Rosenbluth, "Voice Over IP Performance Monitoring," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 2, pp. 9–24, 2001.

[8] M. Ries, O. Nemethova, and M. Rupp, "Video Quality Estimation for Mobile H. 264/AVC Video Streaming," *Journal of Communications*, vol. 3, no. 1, pp. 41–50, 2008.

[9] OpenSignal, "OpenSignal," January 2014. [Online]. Available: http://opensignal.com

[10] Epitiro, "Epitiro," January 2014. [Online]. Available: http://www.epitiro.com

[11] RootMetrics, "RootMetrics," January 2014. [Online]. Available: http://www.rootmetrics.com

[12] S. Sonntag, J. Manner, and L. Schulte, "Netradar - Measuring the wireless world," in *2013 11th International Symposium on Modeling Optimization in Mobile, Ad Hoc Wireless Networks (WiOpt)*, 2013, pp. 29–34.

[13] C. GIST, "Cisco GIST," January 2014. [Online]. Available: http://gistdata.ciscovni.com/

[14] ITU-T, *Methods for Subjective Determination of Transmission Quality*. ITU-T Recommendation P.800, June 1998.

[15] S. Sen, J. Yoon, J. Hare, J. Ormont, and S. Banerjee, "Can They Hear Me Now? A Case for a Client-Assisted Approach to Monitoring Wide-Area Wireless Networks," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011, pp. 99–116.

[16] X. Liu, A. Sridharan, S. Machiraju, M. Seshadri, and H. Zang, "Experiences in a 3G Network: Interplay Between the Wireless Channel and Applications," in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, ser. MobiCom '08. ACM, 2008, pp. 211–222.

[17] J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl, "Anatomizing Application Performance Differences on Smartphones," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '10. ACM, 2010, pp. 165–178.

[18] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "Characterizing Radio Resource Allocation for 3G Networks," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '10. ACM, 2010, pp. 137–150.

[19] M. Jurvansuu, J. Prokkola, M. Hanski, and P. Perala, "HSDPA Performance in Live Networks," in *IEEE International Conference on Communications, 2007. ICC '07*, June 2007, pp. 467–471.

[20] L. Bhebhe and A. Arjona, "Data Outage Across 3G and 2G Wireless Networks," in *WoWMoM 2008. 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks, 2008*, June 2008, pp. 1–8.

[21] A. Gurtov and J. Korhonen, "Effect of Vertical Handovers on Performance of TCP-friendly Rate Control," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 8, no. 3, pp. 73–87, Jul. 2004.

[22] V. Bashko, N. Melnikov, A. Sehgal, and J. Schönwälder, "BonaFide: A Traffic Shaping Detection Tool for Mobile Networks," in *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, Ghent, Belgium, May 2013, pp. 328–335.

[23] ITU-T, *Mean Opinion Score (MOS) Terminology*. ITU-T Recommendation P.800.1, November 2006.

[24] ITU-T, *Subjective Evaluation of Conversational Quality*. ITU-T Recommendation P.805, October 2007.

[25] Hulu, "Hulu Bandwidth Recommendations," January 2014. [Online]. Available: http://www.hulu.com/support/article/197541

[26] Motorola, "Realistic LTE Performance - From Peak Rate to Subscriber Experience," January 2014. [Online]. Available: http://www.motorolasolutions.com/web/Business/_Documents/static%20files/Realistic_LTE_Experience_White_Paper_FINAL.pdf

[27] Alexa, "The Web Information Company," January 2014. [Online]. Available: http://www.alexa.com/

[28] HTTP Archive, "The HTTP Archive - Interesting Stats," January 2014. [Online]. Available: http://www.httparchive.org/

[29] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the Quality of Experience of HTTP Video Streaming," in *2011 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. Dublin, Ireland: IEEE, May 2011, pp. 485–492.

[30] CISCO, "Voice Over IP - Per Call Bandwidth Consumption," January 2014. [Online]. Available: http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a0080094ae2.shtml

[31] R. Kamal, "Mobile Satellite Communication Networks," in *Mobile Satellite Communication Networks In Mobile Computing*. Oxford University Press, 2007, pp. 538–539.

[32] ITU-T, *One-way Transmission Time*. ITU-T Recommendation G.114, November 2003.

[33] ITU-T, *Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 Kbit/s*. ITU-T Recommendation G.723.1, May 2006.

[34] Youtube, "Advanced encoding settings," January 2014. [Online]. Available: https://support.google.com/youtube/answer/1722171?hl=en