RESEARCH

Logistic regression based in-service assessment of mobile web browsing service quality acceptability

Sibila Isak-Zatega[1], Adriana Lipovac[2*] and Vlatko Lipovac[2]

* Correspondence: adriana.lipovac@unidu.hr
[2]University of Dubrovnik, Dubrovnik, Croatia
Full list of author information is available at the end of the article

Abstract

In this paper, we presented a logistic regression model that we applied for assessment of the users' quality of experience with web browsing service over mobile network. With this regard, we chose the Average-Time-to-Connect-TCP network service quality parameter as an independent predictor, obtained by passive monitoring of live traffic data, captured by a passive probe on the mobile network Gn interface, and related to detailed records of the Transport Control Protocol. In parallel with in-service measuring the selected network parameter, we conducted simultaneous subjective tests of the quality of experience acceptability to users, specifically for web browsing service. Particularly, it was found that the model provided correct acceptability classification in 84.5% of cases, while reducing the chosen independent predictor for 100 ms implied increasing the chance of the service acceptability by factor of 1.65. Based on the obtained results, it comes out that the applied logistic regression model provides satisfactory estimation of the web browsing service quality experience acceptability.

Keywords: Quality-of-experience, In-service network monitoring, Regression

1 Introduction

Web browsing is among dominant cellular network applications and is expected to grow by 39% annually over the coming 6 years [1]. Growing users' demand for reliable data delivery comes along with their expectation for adequate Quality-of-Experience (QoE), too, making the latter the most important user decision criterion in selecting a specific service provider. Consequently, network operators are kin to ensure the best possible QoE level, for which the conditio sine qua non is their ability to reliably and accurately assess the achieved customer satisfaction with their services, such as web browsing.

The majority of existing QoE estimation models are based on the mean opinion score (MOS) testing, but MOS-based monitoring of web browsing QoE in particular requires fairly complex metrics [2–4].

Therefore, it has become crucial to relate the subjective QoE to measurable technical service quality parameter(s), which can be in-service monitored in the operator environment, so enabling objective and real-time QoE estimation [3].

Moreover, as network operators are mostly interested in testing users' acceptability of provided web services [5–7], i.e., the "binary measure to locate the threshold of minimum acceptable quality that fulfills user quality expectations and needs for certain application or system" [7], consequently, in recent years, a number of QoE models based on acceptability have been proposed, especially for video signal delivery [8–10], as well as for interactive data services [11].

Specifically, the ITU-T Recommendation G.1030 provides some experimental results regarding the users' perception in relation to web browsing response time, as well as some guidelines for QoE estimation [2]. The according experiments were conducted to evaluate the suitability of the developed network emulator system for the QoE estimation [3] and to validate the ITU-T Recommendation G.1030. The obtained results show logarithmic dependency between the QoE and the page load time for a simple web page.

Furthermore, in contrast to the studies which are mostly based on direct user feedback [2–4], the QoE is sometimes estimated from passive network tests [12, 13]. Specifically, the relationship between the QoE and the Quality-of-Service (QoS) for web browsing services was analyzed based on HyperText Transfer Protocol (HTTP)/Transmission Control Protocol (TCP) traces collected in the network, where cancelation rate of HTTP requests was used for QoE estimation, but without any validation of the achieved results by simultaneous real-life subjective QoE testing.

Moreover, though in some studies subjective user ratings are combined with network-level information, experimental findings coming out of the recorded TCP and HTTP traces and web browsing service QoE are reported only by graphical means and are not backed by any analytical model [14].

Therefore, in this paper, we address the aforementioned challenges by developing the QoE acceptability predictive model for web browsing service over mobile network, where the model is based on network parameters, in-service measurable by passive monitoring of live traffic data, and practically implementable by mobile operators.

As linear regression is not appropriate for modeling acceptability-based QoE, where the outcome variable–the acceptability is binary, logistic regression is the method of our choice. To our best knowledge, so far, no attempt has been made to assess acceptability of the mobile web browsing service by means of the logistic regression.

With this regard, in the previous work [15], the extent of the relationship between the in-service measured live traffic data parameters and the web browsing user QoE in the mobile network was analyzed by using the Spearman's rank-order correlation. Taking into account both strength and direction of the relationship, it came out that the parameters *Average-Time-to-Get-1st-Data* and *Average-Time-to-Connect-TCP* exhibited the strongest relationship with the web browsing QoE evaluated by means of the ordinary 5-point Likert scale (with ratings: excellent, good, fair, poor, bad).

Therefore, in this paper, we followed that indication by applying logistic regression on the selected parameter to assess the users' acceptability of the quality level experienced with web browsing service in particular.

However, in contrast to other investigations using mostly ordinal logistic regression in compliance with the type of test data determining categorical both the independent variables and the dependent one and with numerical MOS rating converted to categorical data, here, we put an accent on today's network operator main QoE imperative with web browsing in particular: to get an objective binary-type customer QoE rating– acceptability, which we model here by binary logistic regression applied to the *Average-Time-to-Connect-TCP* parameter that we found most relevant in this sense.

The rest of this paper is organized as it follows: In Section 2, we review the basics of the logistic regression to be used in the QoE acceptability prediction model. The test setup and tools that we used for conducting the experiment are also described in Section 2, while we present the test results and the analysis of the experimental data in Section 3. Conclusions are drawn in Section 4.

2 Methods

Before analyzing the acquired data by means of logistic regression, we review the concepts of the model and then apply it for the QoE prediction.

2.1 Logistic regression

Regression is mostly used as a means to predict a random variable from a number of mutually independent random variables and a constant.

Specifically, logistic regression is used for predicting the probability that a certain observation will be sorted into one out of two categories of a dichotomous dependent random variable, based on one or more independent random variables, which can be continuous or categorical. In many aspects, logistic regression is similar to linear regression, with the exception of the dependent variable type, which, in contrast to linear regression, does not provide estimated value of the dependent variable, but the probability that it will belong to a certain category, based on the values of the independent variables.

Among the three types of logistic regression, namely binary, ordinal, and nominal, the first one is used when the dependent variable is binary, i.e., takes one out of two categories. Moreover, if a dependent random variable can take more categories, then the ordinal logistic regression or the nominal one is to be used for ordered and unordered categories, respectively.

However, as it is already mentioned in Section 1, though ordinal logistic regression has been most frequently used (even after properly converting MOS scoring), as our focus here is on QoE acceptability, we consider here the binomial logistic regression, commonly referred to simply as logistic regression.

Essentially, it is a supervised machine-learning classification algorithm used to predict the conditional probability:

$$\Pi(x_i) = \Pr(Y = 1/X_i = x_i); \ i = 1, 2, ..., N \tag{1}$$

that a certain individual observation belongs to one out of two categories, i.e., that the corresponding dichotomous dependent random variable Y takes one out of two possible values (1 or 0), conditioned by one or more (N) continuous or categorical mutually independent random variables X_i taking their corresponding values x_i [16–18].

Let us assume the simple linear form of the logit transform (from now on just logit) of $\Pi(x_i)$, specifically for a single value $x_i = x$ [16]:

$$\text{logit}[\Pi(x)] = \ln(\text{odds}) = \ln\left(\frac{\Pi(x)}{1-\Pi(x)}\right) = \alpha + \beta x \qquad (2)$$

where the odds are defined as the ratio of the probability $\Pi(x)$ that the event (outcome of interest) will occur for a particular value x of the random variable X, and the probability $1 - \Pi(x)$ that the event will not occur, while β is the slope coefficient, and the constant α is referred to as the intercept.

From (1) and (2), $\Pi(x)$ can be expressed as:

$$\Pi(x) = \Pr(Y = 1 | X = x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \qquad (3)$$

where the iterative maximum likelihood (ML) method is used for estimating the according α and β values by testing the null hypothesis that these do not make the logistic regression accurate enough. In this case, small significance (represented by the p value) indicates strong evidence to reject the null hypothesis.

In order to use the binomial logistic regression in practice, the following main assumptions need to be fulfilled [18]:

1. Logistic regression requires the observed dependent random variable Y to be dichotomous and a function of one or more mutually independent and non-collinear predicting random variables–predictors.
2. The logit transform must be a linear function of continuous predicting random variables.
3. Each test observation must be independent from others and all test categories should be mutually exclusive and exhaustive.
4. Data must not exhibit significant outliers, high leverage points, or highly influential points; otherwise, the reliability of the estimates may degrade significantly.

2.2 Test setup

The test setup is presented in Fig. 1. As it can be seen, the experiment was carried out on a live network. The test configuration included the client, the gateway that was connected to the live High Speed Packet Access Evolved (HSPA+ Rev.8) mobile network (providing up to 42 Mb/s with 64QAM in downlink, and 11.5 Mb/s with 16QAM in uplink), which is connected to the internet. The gateway ran on Linux OS, while the NetEm [19] enhancement of the Linux traffic control (TC) facilities enabled introducing packet delay and packet loss in the experiment. We chose the test point to be at the Gn interface, where the actually used Oracle Performance Intelligence Center (PIC) [20] with passive probe captured the traffic data, Fig. 1.

Each test participant took part in experiments using the client operating on Windows 8 PC. The client device was connected to the gateway via 100 Mbps Ethernet full duplex link. We used the NetEm network emulator on Ubuntu OS of the gateway to vary the network conditions by adding delay and packet loss. The Huawei E3272 LTE USB modem was used for testing, while being managed by the embedded Connection Manager software, which allowed setting the preferred access network.

Fig. 1 Test system
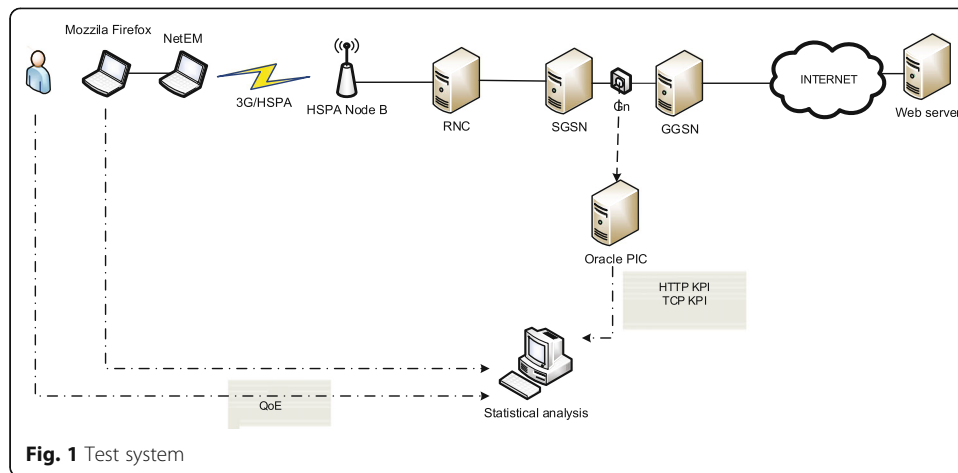
We enabled the HSPA+ to be the preferable access network in the experiment. The client system was connected to internet through the mobile network via the gateway. In both laptops, the automatic software updates were disabled. The participants in the experiment used Mozilla Firefox 35.0.1 web browser. The HTTP and TCP extended Detailed Records (xDR) from the data captured on Gn interface were made available by using ProTrace application on the Oracle PIC platform. This way, we defined and activated new statistical sessions which generated the in-service parameters' values aggregated over 5-min intervals. The parameters were defined from the HTTP and TCP xDR's for the Mobile Station International Subscriber Directory Number (MSISDN) [15] of the test SIM card.

The experiment was conducted by ten users, five female and five male, whose age ranged between 12 and 45 years. All participants used the internet at least 1 h a day and usually via the WiFi access, except when switching to the mobile internet access (only if WiFi was unavailable).

We investigated the relationship between the QoE and in-service parameters through the following test scenario:

Each participant tested web browsing six times under different network conditions, determined by the NetEm (adding delay or packet loss during the experiment).

Duration of a single test was limited to 5 min, while the participants accessed web pages of their choice and simply answered whether the technical quality of web browsing service was acceptable or not, with "yes" or "no", respectively.

Following that, by running statistics sessions in Oracle PIC platform, processing the collected values of the relevant in-service parameter measured on Gn interface–*Average-Time-to-Connect-TCP*, which is the average time between SYN and ACK in the TCP three-way handshake sequence, needed to establish the TCP connections within a 5-min interval [1].

2.3 Test tools
We used the Oracle PIC as monitoring and data gathering system that helps service providers to manage their assets, encompassing network performance, QoS, and

customer analysis [20]. The PIC uses passive probes to capture traffic data and forward probe data units (PDU) to the Integrated xDR Platform (IXP). The IXP stores these traffic data and correlates them into detailed records. The PIC provides applications that mine the detailed records to provide value-added services such as network performance analysis, call tracing, and reporting [21]. For the purpose of this research, we used the HTTP and TCP sessions on the Gn interface of the mobile network, defining parameters, and statistics sessions by using the ProTraq application [21].

Furthermore, we used the NetEm as enhancement of the Linux traffic control facilities that allows adding delay, loss, duplication, and other impairments as well, to the packets outgoing from the selected network interface. NetEm is built using the existing QoS and the differentiated services (DiffServ) facilities in the Linux kernel [19].

3 Discussion and results

We analyzed the fields of data records collected from HTTP and TCP sessions on the Gn interface, and selected the ones to define in-service parameters in Oracle PIC [15]. With this regard, some in-service parameters from HTTP and TCP xDRs, based on the data captured by extensive testing that we made on Gn interface, are presented in the Appendix, while in Fig. 2, the exemplar relevant TCP record time intervals can be seen.

Now, the task is to find out which out of the set of in-service measured parameters, is mostly influencing the QoE acceptability in particular, so to be selected as the logistic regression predicting variable.

With this respect, we consider the correlation to be the best indication, and therefore we calculated it for various parameters, as it is presented in Table 1.

Due to monotonic relationship and evident strong (negative) correlation (Spearman correlation coefficient $r_s = -0.791$, and significance value $p < 0.01$) between the in-service measured parameter *Average-Time-to-Connect-TCP* and users' acceptability of the service quality level [15], we consider the former as the independent predicting random variable X, whose scatter plot with regard to QoE ratings is presented in Fig. 3, while in Table 2, we present its in-service obtained test values selected from the overall data table in the Appendix.

As it can be seen in Fig. 3 and Table 2, just a few sporadic peak values of the Average_Time_to_Connect_TCP were measured (e.g., with just about 10% of them larger than 1.5 s). Observing bottom-up through the protocol stack, various reasons for this could be considered, among them the (eventually) excessive retransmissions of Hybrid Automatic Repeat-reQuest (HARQ) protocol data units (e.g., due to bit errors at the physical layer). These could produce additional delays which propagate upwards the stack causing the TCP 3-way handshaking time-outs (such as e.g., retransmission time-out (RTO)), which imply even further delays of the TCP connection setup time.

3.1 Verifying the logistic regression assumptions

The first assumption for the logistic regression can be considered holding in this case, as it can be seen in Table 2 that the observed dependent random variable Y is obviously dichotomous and a function of just a single predicting continuous random variable (implying that, in this case, multicollinearity among the predictors is not an issue).
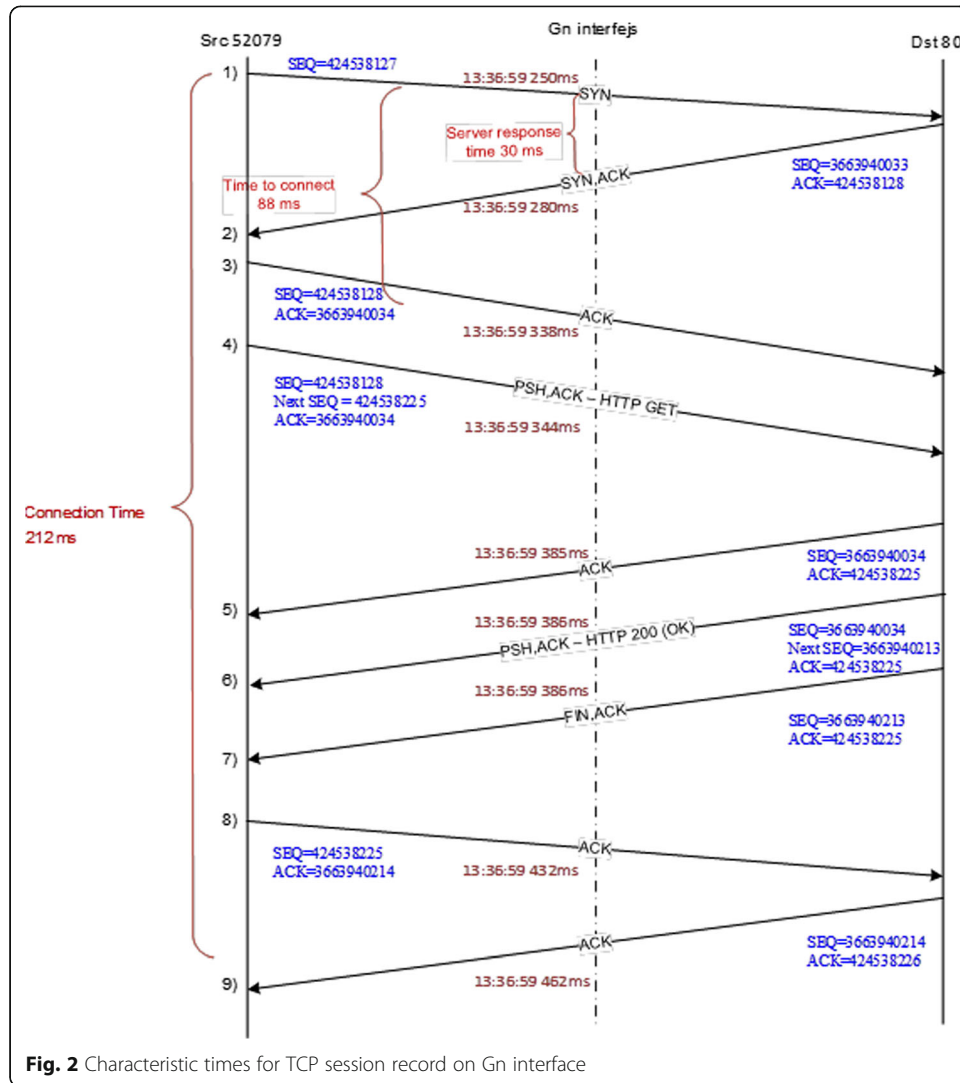
Fig. 2 Characteristic times for TCP session record on Gn interface

Regarding the second assumption, we used the Box-Tidwell (1962) procedure [22] to test whether the logit transform is a linear function of the predictor, effectively by adding the non-liner transform $X{\cdot}\ln X$ of the original predictor X, as a second, so-called interaction variable, and testing the null-hypothesis that adding it made no better prediction. As it can be seen in the according table that is presented in Section 3.2, we found the logit linearity condition holding, with just minor non-linearity possible.

Further on complying with the third assumption, we did each test independently of others, with all test categories being mutually exclusive and exhaustive.

Moreover, the data exhibited quite balanced behavior with no significant outliers and no leverage or influential points.

Consequently, we can justifiably expect that the conducted logistic regression procedure finally provided valid conditional probability $\Pi(x)$ that the dependent random variable Y takes one out of two possible values (1 or 0), conditioned by a single (in our case) predicting random variable X taking the value x.

Table 1 Spearman correlation coefficient between QoE and in-service tested parameters [15]

In-service parameter	QoE		
	Likert 5	Likert 3	Binary
Succ_rate_HTTP	0.168	0.149	0.309*
Cancellation_rate_HTTP	− 0.320*	− 0.301*	− 0.406**
Avg_Server_Response_Time_HTTP	− 0.114	− 0.179	− 0.130
Avg_Time_To_Get_1st_Data_All_HTTP	− 0.790**	− 0.792**	− 0.647**
Retransmit_DL_Vol_Ratio_HTTP	− 0.201	− 0.169	− 0.236
Retransmit_UL_Vol_Ratio_HTTP	− 0.224	− 0.230	− 0.224
Max_DL_throughput_HTTP	0.616**	0.592**	0.515**
Max_UL_throughput_HTTP	0.575**	0.554**	0.488**
Avg_Time_To_Get_Data_HTTP	0.034	− 0.010	− 0.075
Avg_Transaction_Time_HTTP	− 0.510**	− 0.497**	− 0.494**
Avg_Transfer_Time_HTTP	− 0.491**	− 0.487**	− 0.523**
Radio_TCP_succ	0.239	0.241	0.147
Avg_Server_Response_Time_TCP,	0.066	− 0.029	0.008
Avg_Time_to_Connect_TCP	− 0.831**	− 0.835**	− 0.669**
Avg_DL_throughput_TCP	0.482**	0.497**	0.402**
Avg_UL_throughput_TCP	0.652**	0.624**	0.534**
Ratio_DL_Bytes_Retr_TCP	− 0.217	− 0.181	− 0.210
Ratio_UL_Bytes_Retr_TCP	− 0.199	− 0.158	− 0.235
Ratio_DL_Packets_Retr_TCP	− 0.304*	− 0.307*	− 0.303*
Ratio_UL_Packets_Retr_TCP	− 0.133	− 0.115	− 0.175
Avg_DL_Max_RTT_TCP	− 0.506**	− 0.483**	− 0.423**
Avg_UL_Max_RTT	0.129	0.079	0.085
Avg_Duration_Connection	− 0.150	− 0.125	− 0.194

**Correlation is significant at the 0.01 level (2-tailed)
*Correlation is significant at the 0.05 level (2-tailed)

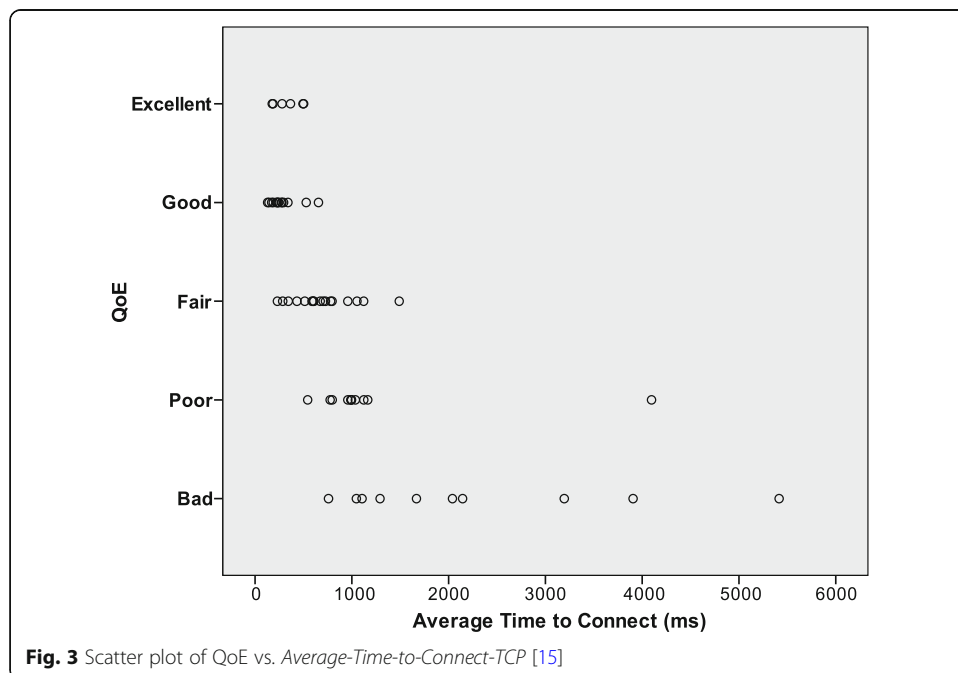**Fig. 3** Scatter plot of QoE vs. *Average-Time-to-Connect-TCP* [15]

Table 2 Independent predicting variable *Average-Time-to-Connect-TCP* (ms), in-service measured at the Gn interface

No.	Avg_Time_to_Connect_TCP	No.	Avg_Time_to_Connect_TCP	No.	Avg_Time_to_Connect_TCP	No.	Avg_Time_to_Connect_TCP
1	1123	16	601	31	4097	46	496
2	172	17	1054	32	222	47	295
3	502	18	2145	33	656	48	274
4	367	19	514	34	177	49	1668
5	433	20	798	35	344	50	544
6	589	21	776	36		51	704
7	608	22	989	37	340	52	759
8	781	23	959	38	5415	53	959
9	1490	24	1047	39	529	54	1292
10	997	25	286	40	674	55	2040
11	3195	26	189	41	728	56	1106
12	1165	27	231	42	1037	57	278
13	186	28	238	43	995	58	144
14	130	29	229	44	3906	59	280
15	246	30	797	45	1124		

3.2 Test cases and estimated logistic regression parameters

We consider a case (sample) to be a repeatable single test made by a single participant. A number of recommended values exists for the required minimum number of samples (cases) ranging from 15 to 50, but we adopted 60 samples per independent random variable [18], as the ML-based logistic regression estimation significantly degrades for rare test cases.

So, the counts of cases included/missing in the analysis are given in Table 3 (in accordance with Table 2), while Table 4 presents how the outcome random variable Y is encoded.

The logistic regression coefficients for the model with independent random variable *Average-Time-to-Connect-TCP* are estimated to take the values of $\alpha = 4.746$, $\beta = -0.005$, while their properties—the standard error (S.E.), the Wald Chi-square (χ^2) test value [18] for D.F. degrees of freedom, and the significance expressed by the p value—are presented in Tables 5 and 6.

So, as we can see from the above tables, the Wald test [20] evaluates the independent random variable *Average-Time-to-Connect-TCP* as statistically significant in the model, as the p value is found to be very low: $p < 0.001$.

Table 3 Independent random variable X test cases

Unweighted cases		N	Percent
Selected cases	Included in analysis	58	98.3
	Missing cases	1	1.7
	Total cases	59	100.0
Unselected cases		0	0
Total cases		59	100.0

Moreover, as it is mentioned in Section 2, we tested the linearity assumption determining the validity of logistic regression, by applying the Box-Tidwell (1962) procedure [22]. Accordingly, we tested the null-hypothesis that adding the new variable:

$$Avg\_Time\_To\_Connect\_TCP \times \ \ln(Avg\_Time\_To\_Connect\_TCP)$$

into regression would make no better prediction.

As it can be seen from low p values in Table 7, we found the logit linearity condition holding, with just minor non-linearity possible.

3.3 Intercept-only model and its extension by prediction

Back to (2), at first, let us consider the model without taking into account the independent random variable *Average-Time-to-Connect-TCP*, i.e., when:

$$\Pi(x) = \ \Pr(Y = 1|X = x) = \ \Pr(Y = 1) = \Pi \tag{4}$$

which effectively modifies (2) into:

$$\mathrm{logit}\Pi = \ \ln(odds) = \ \ln\left(\frac{\Pi}{1-\Pi}\right) = \alpha \tag{5}$$

Accordingly, in the next two tables, the outputs related to the model that includes only the intercept value α, are presented. Such incomplete model predictions depend purely on what category occurred most frequently in the data set, in accordance with (4)/(5). It simply predicts that the service is acceptable, as the majority of participants in the experiment considered the service acceptable (38 out of 58 participants answered "yes"). So, applying this "best guess" strategy, one would be right for 65.5% of time (Table 8).

Accordingly, the estimated statistics for this special case is presented in Table 9.

As exponential of α given in Table 9, the odds are estimated to be equal to 1.9, which conforms to the ratio 38/20 of the counts of users who found the service acceptable and not acceptable, respectively.

Now let us turn to the analysis of logistic regression with included independent random variable X, i.e., the *Average-Time-to-Connect-TCP* as a predictor.

The likelihood ratio (LR) test is used to judge the null hypothesis that including the *Average-Time-to-Connect-TCP* random variable into the model does not significantly increase the ability to predict the decisions made by the subjects. This essentially implies testing the ratio:

$$G = -2 \ \ln\left(\frac{L0}{L1}\right) \tag{6}$$

of likelihoods $L0$ and $L1$ of test data representing the zero-valued and the maximum likelihood estimate (MLE) of the parameter of interest, respectively. Under the

Table 4 Dependent random variable Y encoding

Original value	Coded value
No	0
Yes	1

Table 5 Estimated logistic regression intercept and its properties

Intercept a	S.E.	Wald	D.F.	p value
4.764	1.261	14.261	1	0.00016

hypothesis that $\beta = 0$, the statistics G follows the Chi-square distribution with 1 degree of freedom [17].

The according test results are presented in Table 10.

From Table 10, it can be seen that, for the Chi-square model with 1 degree of freedom and the value of 35.817, it comes out that $p < 0.00001$, and we justifiably reject the null hypothesis. So the results of this test indicate that including *Average-Time-to-Connect-TCP* random variable into the model statistically significantly increases the ability to predict the acceptability of the service to the users.

3.4 Testing the logistic regression model goodness of fit

Furthermore, adequacy of the model can be assessed by means of the Hosmer and Lemeshow goodness-of-fit test, which actually evaluates inadequacy of the model in predicting categorical outcomes, i.e., the hypothesis that the observed data are significantly different from the predicted values coming out of the model.

The test essentially partitions n observations into g approximately equal-size groups–deciles, so that the first group contains approximately $n/10$ observations with the smallest estimated probability, and the last group of approximately $n/10$ observations with the largest estimated probabilities [17].

The statistics is:

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_{1k} - E_{1k})^2}{E_{1k}(1 - \xi_k)} ; E_{1k} = s_k \xi_k \tag{7}$$

where O_{1k} is the count of observations with $Y = 1$ (out of s_k observations in total) in the kth group, and E_{1k} is the expected count of the event in the kth group, whereas ξ_k is the average predicted event probability for the kth group.

The statistics of (7) is close to χ^2 distribution with 8 degrees of freedom (as for totally $g = 10$ groups, it is: $10 - 2 = 8$). Small enough p value (< 0.05) implies that the model poorly fits to data.

The $2 \times g$ contingency table presents the observed and the expected counts of the event $Y = 1$. Accordingly, resulting from our test data are entries in Table 11.

Finally, as it can be seen in Table 12, the obtained high p value indicates low-significance inadequacy of the fitting model, which implies that the model is not to be considered inadequate (but opposite, i.e., adequate).

3.5 Category prediction

Furthermore, as the logistic regression estimates the probability of the event that the service is acceptable to users, we adopt a typical decision threshold of 0.5, meaning that

Table 6 Estimated logistic regression slope and its properties

Predictor variable	Slope β	S.E.	Wald	D.F.	p value
Average-Time-to-Connect-TCP	− 0.005	0.001	12.116	1	0.0005

Table 7 Testing linearity assumption

Predicting variable	B	S.E.	Wald	D.F.	p value
Avg_Time_to_Connect_TCP	− 0.034	0.014	6.247	1	0.012
Avg_Time_To_Connect_TCP × ln(*Avg_Time_To_Connect_TCP*)	0.004	0.002	5.434	1	0.020
Intercept *a*	A	S.E.	Wald	D.F.	p value
	8.014	2.545	9.914	1	0.002

if the estimated probability is greater than or equal to 0.5, the event is classified as the one that will happen; otherwise, the event is classified as the one that will not happen [23].

Accordingly, the observed and predicted classifications are presented in Table 13.

As it is already pointed out in Section 3.3, Table 8, where the classification includes just the intercept constant, we can see that 65.5% of cases overall could be correctly classified by simply considering all cases (to be classified) as choosing "yes" for acceptability.

However, with the independent random variable included in the model, the so-called Percentage-Accuracy-in-Classification (PAC) [18] can be seen in Table 13 to be equal to 84.5%, as the model correctly classified that many cases on a relative scale, which is a significant improvement with regard to the case of classification without the predictor variable.

Another classification feature is the sensitivity, which is the percentage of cases with the target category correctly predicted by the model when the quality of services was evaluated as acceptable ("yes"). So, as it is presented in Table 13, 86.8% of test cases, when participants rated service as acceptable, were also classified by the model as acceptable.

On contrary, the specificity is the percentage of cases that were found to not have the target category [21], i.e., which were correctly classified by the model when the service was not rated as acceptable. In our study, the specificity was found to be equal to 80.0%, meaning that 80% of participants who did not rate service as acceptable were correctly classified by the model, Table 13.

The positive predictive value is the percentage of correctly classified cases exhibiting the target characteristic, relative to the total count of cases predicted to have the target characteristic. In this case, simple calculus provides:

$$33/(33 + 4) = 89.19\%$$

Table 8 Classification without the independent predictor

Observed		Predicted		Correct prediction %
		Acceptability		
		No	Yes	
Acceptability	No	0	20	0.0
	Yes	0	38	100.0
Overall prediction %				65.5

Table 9 The intercept-only model attributes

Intercept a	S.E.	Wald	D.F.	p value
0.642	0.276	5.398	1	0.020

meaning that, out of all cases predicting the service acceptable, for 89.19% of them, the prediction is correct.

The negative predictive value is the percentage of correctly classified cases without the target characteristics, relative to the total count of cases predicted as not having the target characteristics. In this case, it is:

$$16/(16+5) = 76.19\%,$$

meaning that out of all cases predicting the service not acceptable, for 76.19% of them, the prediction is correct.

Finally, by substituting the values of α = 4.746, β = − 0.005 from Section 3.2, into the regression Eq. (2), the latter can be rewritten as:

$$\text{logit}[\Pi(x)] = \ln(odds) = 4.764 - 0.005x \tag{8}$$

where x = *Average-Time-to-Connect-TCP*, while (3) turns into:

$$\Pi(x) = \frac{e^{4.764-0.005x}}{1 + e^{4.764-0.005x}} \tag{9}$$

The conditional probability (x) that, having measured *Average-Time-to-Connect-TCP* milliseconds, the acceptable quality of web browsing service will result is the target logistic regression test outcome, which is plotted in Fig. 4.

As it can be seen on the above graph, the transition between the acceptability and non-acceptability is rather steep, as the curve exhibits a threshold effect around the predictor value of 1 s.

As an example, let us calculate how much would the odds be affected by reducing the *Average-Time-to-Connect-TCP* parameter for 100 ms. With this regard, we simply substitute the increment as follows:

$$\ln[odds(x + 100)] = 4.764 - 0.005 \cdot (x + 100) \tag{10}$$

$$\ln[odds(x)] - \ln[odds(x + 100)] = 0.5 \tag{11}$$

$$\frac{odds(x)}{odds(x + 100)} = e^{0.5} = 1.65 \tag{12}$$

According to (12), it comes out that by reducing the *Average-Time-to-Connect-TCP* parameter for 100 ms, the chance of success, i.e., the chance that the service is acceptable, increases by the factor of 1.65.

Table 10 Likelihood ratio test statistics

Chi-square	D.F.	p value
35.817	1	<0.00001

Table 11 Contingency table for Hosmer and Lemeshow test

K	Acceptability					Total
	No		Yes			
	Observed	Expected	Observed	Expected		
1	6	5.993	0	0.007		6
2	3	4.897	3	1.103		6
3	5	3.630	1	2.370		6
4	3	2.486	3	3.514		6
5	3	1.418	3	4.582		6
6	0	0.759	6	5.241		6
7	0	0.397	6	5.603		6
8	0	0.198	6	5.802		6
9	0	0.148	6	5.852		6
10	0	0.073	4	3.927		4

Table 12 Hosmer and Lemeshow test

Chi-square	D.F.	*p* value
9.531	8	0.300

Table 13 Classification; observed vs. predicted

Observed		Predicted		Percentage of correct
		Acceptability		
		No	Yes	
Acceptability	No	16	4	80.0
	Yes	5	33	86.8
Overall percentage				84.5

Fig. 4 Probability   (x) of acceptable web service, conditioned by *Average-Time-to-Connect-TCP*
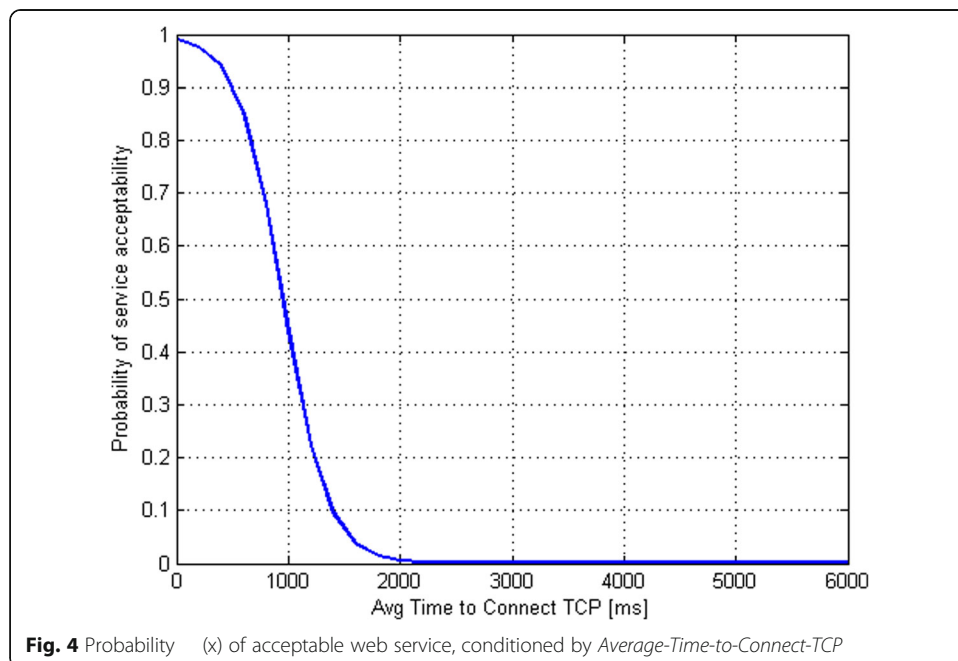
4 Conclusions

We proposed a simple logistic regression model with a single independent predicting variable, namely the *Average-Time-to-Connect-TCP* network parameter, derived from live traffic data captured by passive probe on the Gn interface of the mobile network, to estimate the users' quality of experience acceptability of the web browsing service.

In parallel, we conducted simultaneous subjective users' service quality acceptability tests with a number of participants, to finally correlate the obtained values to detailed records of the TCP protocol.

The model was found to provide correct estimation of the experienced service quality acceptability, with high statistical significance determined by Chi-squared value above 35, p value below 0.0005, and correct classification in 84.5% of cases.

More specifically, the sensitivity and specificity were found to be equal to 86.8% and 80%, respectively, while the positive and negative prediction values were evaluated to be equal to 89.19% and 76.19%, respectively.

Reducing by 100 ms the network service parameter that is selected as the predicting variable was found to increase the chance of the service acceptability by the factor of 1.65.

We plan to extend the proposed approach application range and enhance its ability to predict real-life acceptability of service quality experience, by involving more experimental scenarios and wide-area users, as well as other aspects such as context and extended set of parameters. Moreover, the full-scale measurement campaign would go beyond resource-limited preliminary tests reported here, and in 4G/5G environment, where this approach and analysis still apply.

Table 14 In-service measured test data (time units, ms)

Nb	Succ_rate_HTTP	Cancellation_rate_HTTP	Avg_Server_Response_Time_HTTP	Avg_Time_To_Get_1st_Data_All_HTTP	Retransmit_DL_Vol_Ratio_HTTP	Retransmit_UL_Vol_Ratio_HTTP	Max_DL_throughput_HTTP	Max_UL_throughput_HTTP	Avg_Time_To_Get_Data_HTTP	Avg_Transaction_Time_HTTP	Avg_Transfer_Time_HTTP
1	0.84	0.04	51	721	0.004	0.002	987.75	0.11	61	29581	327
2	0.73	0.11	36	76	0.004	0.002	1156.56	0.07	111	21737	228
3	0.73	0.05	161	160	0.016	0.002	562.06	0.09	228	9424	476
4	0.82	0.01	42	117	0.003	0.003	919.9	0.13	78	8310	170
5	0.91	0	65	156	0.001	0.002	1338.15	0.14	73	6862	225
6	0.78	0	42	316	0	0.002	897.33	0.04	52	27529	323
7	0.86	0.01	29	123	0.001	0.002	595.46	0.02	33	12140	370
8	0.93	0.02	57	479	0.007	0.001	717.09	0.11	448	21879	742
9	0.56	0.34	125	601	0.011	0.002	475.54	0.04	510	27149	1671
13	0.92	0.03	71	121	0.028	0.001	757.79	0.13	768	19490	959
14	0.9	0	37	62	0.001	0.001	1411.36	3.35	47	18136	150
15	0.72	0.12	52	65	0.01	0.002	1160.95	0.04	466	17095	721
16	0.81	0.07	111	412	0.022	0.001	351.83	3.31	635	22105	1479
17	0.9	0	67	535	0.042	0.002	245.6	0.26	82	20816	271
19	0.94	0	83	587	0.001	0.002	116.66	0.07	93	8656	478
25	0.93	0	61	143	0.115	0.002	576.08	0.11	97	13517	501
26	0.91	0	35	109	0.1	0.002	66.17	0.15	54	10342	71
27	0.95	0.01	71	89	0.047	0.002	266.12	0.23	81	17150	278
28	0.93	0	46	109	0.137	0.002	189.68	0.73	62	13525	179
29	0.9	0	34	85	0.014	0.002	490.61	0.11	40	6520	144
30	0.88	0.03	69	417	0.018	0.002	257.81	0.12	93	12987	421
32	0.96	0	61	159	0.003	0.001	534.64	0.36	89	15170	170
33	0.91	0.01	39	127	0.007	0.001	447.43	1.28	66	6870	339

Table 14 In-service measured test data (time units, ms) *(Continued)*

Nb	Succ_rate_HTTP	Cancellation_rate_HTTP	Avg_Server_Response_Time_HTTP	Avg_Time_To_Get_1st_Data_All_HTTP	Retransmit_DL_Vol_Ratio_HTTP	Retransmit_UL_Vol_Ratio_HTTP	Max_DL_throughput_HTTP	Max_UL_throughput_HTTP	Avg_Time_To_Get_Data_HTTP	Avg_Transaction_Time_HTTP	Avg_Transfer_Time_HTTP
34	0.92	0.02	64	74	0.004	0.002	886.49	0.27	252	11613	503
35	0.78	0.09	158	289	0.013	0.002	881.13	0.07	823	18985	1433
36		0				0	0	0			
37	0.86	0.04	91	202	0.002	0.001	774.61	3.63	251	18578	782
38	0.53	0.27	30	4274	0.201	0.003	39.49	0.04	31	51565	10318
39	0.86	0.05	50	249	0.01	0.002	751.9	0.15	555	15677	1031
40	0.83	0.1	191	193	0.006	0.002	759.6	0.21	1082	12033	1944
41	0.91	0.03	50	166	0.01	0.002	784.5	0.1	345	9406	793
45	0.96	0.01	46	398	0.002	0.002	663.3	0.04	58	14220	418
46	0.93	0	33	147	0.027	0.002	1153.59	0.14	34	17632	453
47	0.84	0.03	53	50	0.002	0.001	1075.61	38.62	89	3630	224
48	0.95	0.01	67	113	0.006	0.002	829.62	0.34	170	12694	286
50	0.92	0.02	110	234	0.007	0.002	266.43	0.15	83	13349	190
53	0.89	0.09	67	351	0	0.002	317.41	0.04	95	13554	1784
57	0.7	0.19	49	135	0.022	0.002	1291.01	0.28	77	21584	736
58	0.88	0	46	75	0.009	0.002	878.76	0.15	61	9669	253
59	0.89	0.03	55	197	0.003	0.003	330.41	0.25	70	18153	206

Table 15 In-service measured test data (time units, ms)

Nb	Radio_TCP_succ	Avg_Server_Response_Time_TCP	Avg_Time_to_Connect_TCP	Avg_DL_throughput_TCP	Avg_UL_throughput_TCP	Ratio_DL_Bytes_Retr_TCP	Ratio_UL_Bytes_Retr_TCP	Ratio_DL_Packets_Retr_TCP	Ratio_UL_Packets_Retr_TCP	Avg_DL_Max_RTT_TCP	Avg_UL_Max_RTT	Avg_Duration_Connection
1	0.866	43	1123	14306	566	0.01	0.055	0.044	0.099	1892	202	20092
2	0.892	44	172	20163	768	0.001	0.029	0.017	0.067	725	129	10034
3	0.925	280	502	1779	126	0.014	0.051	0.072	0.139	2407	1206	41677
4	0.9	71	367	1872	215	0.001	0.005	0.017	0.02	1316	2556	17342
5	0.918	62	433	2064	268	0.003	0.025	0.028	0.081	1032	815	18727
6	0.892	48	589	9276	464	0.001	0.028	0.028	0.095	1277	741	23832
7	0.935	48	608	1348	127	0.004	0.039	0.072	0.109	1658	719	25966
8	0.888	75	781	644	262	0.014	0.034	0.093	0.124	1523	1077	21706
9	0.834	36	1490	5133	291	0.013	0.053	0.054	0.081	1549	437	17586
13	0.942	95	186	1152	729	0.027	0.041	0.087	0.157	807	220	20785
14	0.921	43	130	21855	765	0.005	0.036	0.021	0.09	932	839	15410
15	0.852	38	246	18705	738	0.012	0.035	0.025	0.068	1515	917	15940
16	0.839	64	601	450	134	0.017	0.094	0.123	0.173	2421	1070	42568
17	0.832	109	1054	436	113	0.049	0.136	0.2	0.248	1508	1057	35395
19	0.917	63	514	1791	128	0.015	0.043	0.067	0.1	2054	630	33018
25	0.939	68	286	901	148	0.062	0.091	0.12	0.207	1991	992	33420
26	0.943	57	189	1630	153	0.035	0.064	0.057	0.154	1955	342	34860
27	0.916	57	231	6179	395	0.058	0.103	0.087	0.253	1273	527	39622
28	0.963	65	238	2297	240	0.1	0.104	0.15	0.243	671	2882	44288
29	0.923	54	229	5610	223	0.004	0.083	0.048	0.13	2604	853	43122
30	0.928	97	797	663	97	0.01	0.1	0.104	0.16	4070	2352	36399
32	0.954	69	222	826	307	0.007	0.038	0.065	0.094	985	588	22124
33	0.867	63	656	687	193	0.021	0.116	0.124	0.188	1955	568	32154

Table 15 In-service measured test data (time units, ms) (Continued)

Nb	Radio_TCP_succ	Avg_Server_Response_Time_TCP	Avg_Time_to_Connect_TCP	Avg_DL_throughput_TCP	Avg_UL_throughput_TCP	Ratio_DL_Bytes_Retr_TCP	Ratio_UL_Bytes_Retr_TCP	Ratio_DL_Packets_Retr_TCP	Ratio_UL_Packets_Retr_TCP	Avg_DL_Max_RTT_TCP	Avg_UL_Max_RTT	Avg_Duration_Connection
34	0.926	76	177	1471	416	0.011	0.048	0.104	0.168	1087	333	26770
35	0.83	68	344	853	404	0.011	0.049	0.053	0.102	1383	1136	17966
36	0											
37	0.859	68	340	1082	193	0.003	0.04	0.062	0.126	1663	877	28285
38	0.851	38	5415	401	65	0.205	0.372	0.335	0.351	5201	42	33444
39	0.857	55	529	830	219	0.014	0.041	0.085	0.143	1204	564	25800
40	0.862	100	674	1028	129	0.008	0.038	0.061	0.129	3084	1732	32403
41	0.925	54	728	731	90	0.014	0.044	0.086	0.132	3058	1818	36531
45	0.888	72	1124	852	120	0.008	0.044	0.056	0.126	1915	385	25460
46	0.938	39	496	10983	423	0.023	0.125	0.048	0.205	1388	865	25776
47	0.92	70	295	3497	678	0.094	0.045	0.048	0.123	804	1496	36630
48	0.94	74	274	1430	170	0.01	0.068	0.109	0.174	1715	1118	37727
50	0.894	75	544	1059	198	0.006	0.018	0.033	0.043	1216	854	15009
53	0.914	36	959	2303	115	0.002	0.023	0.039	0.035	3720	3343	33860
57	0.882	66	278	1762	175	0.029	0.084	0.095	0.215	3522	1117	42962
58	0.925	53	144	5228	222	0.003	0.038	0.013	0.102	742	149	15005
59	0.947	67	280	1729	187	0.005	0.049	0.082	0.111	1110	615	24777

5 Appendix

Abbreviations

DiffServ: Differentiated services; HSPA+: High Speed Packet Access Evolved; HTTP: Hypertext Transfer Protocol; IXP: Integrated xDR Platform; LR: Likelihood ratio; ML: Maximum likelihood; MLE: Maximum likelihood estimate; MOS: Mean opinion score; MSISDN: Mobile Station International Subscriber Directory Number; OS: Operating system; PAC: Percentage-Accuracy-in-Classification; PDU: Probe data units; PIC: Performance Intelligence Center; PC: Personal computer; QoE: Quality-of-Experience; QoS: Quality-of-Service; TC: Traffic control; TCP: Transmission Control Protocol; xDR: Extended Detailed Records

Authors' contributions

SIZ made in-service passive monitoring of traffic data on the mobile network Gn interface, as well as simultaneous tests of the web browsing service quality acceptability to users, which define independent and dependent variables of the model, respectively. AL finalized the logistic regression model, to accommodate testing the network service quality acceptability based on the chosen particular network parameter, and the text. VL set up general guidelines for this work, specifically with regard to the analysis of test results and model verification. The author(s) read and approved the final manuscript.

Author details

[1]BH Telecom, Sarajevo, Bosnia and Herzegovina. [2]University of Dubrovnik, Dubrovnik, Croatia.

References

1.   Ericsson mobility report on the pulse of the networked society, November 2016
2.   ITU-T Recommendation G.1030, Estimating end-to-end performance in IP networks for data applications, 2005
3.   E. Ibarrola, F. Liberal, I. Taboada and R. Ortega, "Web QoE evaluation in multi-agent networks: validation of ITU-T G.1030," *2009 Fifth International Conference on Autonomic and Autonomous Systems*, Valencia, 2009, pp. 289-294.
4.   P. Reichl, B. Tuffin, R. Schatz, Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. Telecommun. Syst. **52**(2), 587–600 (2013)
5.   International Telecommunication Union, "Vocabulary and effects of transmission parameters on customer opinion of transmission quality, amendment 2", ITU-T Recommendation P10/G.100, 2006
6.   P. Spachos, W. Li, M. Chignell, A. Leon-Garcia, L. Zucherman and J. Jiang, "Acceptability and quality of experience in over the top video," *2015 IEEE International Conference on Communication Workshop (ICCW)*, London, 2015, pp. 1693-1698.
7.   Ernst Biersack, Christian Callegari, Maja Matijašević, "Data Traffic Monitoring and Analysis", Vol. 7754, Ed. 1, Springer-Verlag Berlin Heidelberg, 2013
8.   Song, Wei & Tjondronegoro, Dian & Himawan, Ivan, "Acceptability-based QoE management for user-centric mobile video delivery: a field study evaluation," *MM* 2014 – *Proceedings of the 2014 ACM Conference on Multimedia*. doi: https://doi.org/10.1145/2647868.2654923.
9.   Agboma, F. and Liotta, A., "Quality of experience management in mobile content delivery systems", Telecommun Syst (2012), 49, 1, pp. 85-98.
10. Song, W. and Tjondronegoro, D., "Acceptability-based QoE models for mobile video", IEEE T. Multimedia (2014), 16, 3, pp. 738 – 750.
11. R. Schatz, S. Egger and A. Platzer, "Poor, good enough or even better? bridging the gap between acceptability and QoE of mobile broadband data services," *2011 IEEE International Conference on Communications (ICC)*, Kyoto, 2011, pp. 1-6.
12. Star Khirman and Peter Henriksen, "Relationship between Quality-of-Service and Quality-of-Experience for public internet service", Passive and Active Network Measurement workshop, March 2002
13. Collange and J. L. Costeux, Passive estimation of quality of experience. J. Univ. Comput. Sci. **14**(5), 625–641 (2008)
14. Raimund Schatz and Sebastian Egger, "Vienna surfing – assessing mobile broadband quality in field", Proceedings of the 1st ACM SIGCOMM Workshop on Measurements Up the Stack (W-MUST). ACM, 2011
15. S. Isak-Zatega and V. Lipovac, "In-service assessment of mobile services QoE from network parameters," 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 2016, pp. 1-7.
16. Peng, Joanne & So, Tak-Shing, "Logistic regression analysis and reporting: a primer", Understanding statistics: statistical issues in psychology. Education., 2002, pp. 31-70.
17. David W. Hosmer, Stanley Lemeshow, "Applied Logistic Regression", Second Edition, A Wiley-Interscience Publication, John Wiley & Sons, Inc., 2000

18. Laerd Statistics, "Binomial logistic regression using SPSS Statistics", Statistical tutorials and software guides., 2005, retrieved from https://statistics.laerd.com/
19. Stephen Hemminger, "Network emulation with NetEm", Linux Conf Au, 2005
20. Oracle and/or its affiliates, Oracle Communication Performance Intelligent Center, Oracle data sheet, 2013 K. Elissa, unpublished.
21. Oracle, Oracle® Communications Performance Intelligence Center ProTrace User's Guide, Release 10.1.5, E56987 Revision 1, 2015
22. G.E.P. Box, P.W. Tidwell, Transformation of the independent variables. *Technometrics* **4**, 531–550 (1962)
23. Fox, J., "Applied Regression, Linear Models, and Related Methods", SAGE Publications, 1997

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.