QoE Analysis of the Setup of Different Internet Services for FIFO Server Systems

Tobias Hoßfeld[1], Martín Varela[2], Poul E. Heegaard[3], and Lea Skorin-Kapov[4]

[1] University of Duisburg-Essen, Modeling of Adaptive Systems,
Essen, Germany, tobias.hossfeld@uni-due.de
[2] Independent Researcher
Oulu, Finland, martin@varela.fi
[3] Department of Telematics, NTNU, Norwegian University of Science and Technology,
Trondheim, Norway, poul.heegaard@item.ntnu.no
[4] University of Zagreb, Department of Electrical Engineering and Computing,
Zagreb, Croatia, Lea.Skorin-Kapov@fer.hr

Abstract Queueing systems following a first-in-first-out (FIFO) strategy are well understood and various results are known for the response time of the system. However, the question arises how the results look like when taking into account a user-centric point of view. To this end, an M/M/1-FIFO system is investigated for the setup of different Internet services. In this tutorial paper, the impact of the system's delay on Quality of Experience (QoE) is considered for 1) YouTube video (initial playout delay), 2) authentication in social networks, 3) wireless 3G Internet connection setup. Existing QoE models are used to map the response time in the system, corresponding to the waiting time for users until the service is setup, to Mean Opinion Scores (MOS) as a measure of QoE. The system is then evaluated in terms of overall QoE and QoE fairness for the three services considered, under different load scenarios. The results show how different such systems and response times are perceived by users of different services. Further, the dimensioning of FIFO systems with respect to QoE only requires us to consider the overall QoE.

1 Introduction

Quality of Experience is the "degree of delight or annoyance of the user of an application or service" [2]. Furthermore, "it results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state". In real services, the quality perceived by the user is heavily affected by the performance of the underlying system, and in particular, of the network. In general, the factors influencing QoE can be classified into Human, System and Context influence factors (IFs) [2]. In practical applications, the Human IFs are hard (if even possible) to measure and affect, and often times, the Context factors are similarly intractable (though some Context-related factors can be measured and taken into account e.g., in QoE models).

System IFs, in contrast, are both better understood and somewhat possible to control. We can further refine System IFs into Application and Resource IFs, as described in [17]. Application factors can relate to e.g., choice of encoding, use of error concealment / correction mechanisms. Resource factors relate to, for example, device capabilities, network resources and state, etc.

In this tutorial paper we focus on the analysis of systems from a QoE perspective, and in particular, on the network performance. We illustrate the approach through the use of simple FIFO queues. This type of approach has been used with good results for instance for analysing the effects of Forward Error Correction (FEC) on VoIP streams [16,1]. Of course, there are more realistic and complex models, but the core message of the paper can be best explained with an analytically simple queuing system, without loss of generality.

Based on existing studies [6], QoE models for the following Internet services are utilized by mapping response times to Mean Opinion Score (MOS) values, 1) YouTube until the video playout starts, 2) authentication in social networks, 3) wireless 3G Internet connection setup. Once again, the use of the MOS provides the simplest possible way to perform a QoE analysis, despite it being sub-optimal for e.g., control or business purposes [7].

The remainder of the paper is structured as follows. Section 2 revisits existing results on the performance of the M/M/1-FIFO queue. In particular, the response time distribution is available. Section 3 provides a background on existing QoE models for waiting times and introduces the mapping functions used in the study of the FIFO system. In addition, the QoE fairness metric and its computation is discussed. Section 4 shows the methodology to analytically and numerically derive the QoE results. The numerical results for the different services are analyzed in Section 5. Section 6 concludes this work with an outlook on future work.

2 Performance of the M/M/1-FIFO Queue

The setup of the different Internet services is modeled as M/M/1-FIFO queue. The different user requests arrive in the system and are served in a first-in-first-out manner, i.e. in the order of user arrival. Since there is only one server, users may have to wait until they are served. The total response time of the system (also called sojourn time) includes the waiting time and the processing time of a user request (also called a job in queueing theory). The M/M/1 system is well investigated, see [10] or e.g., [11,5] for more recent textbooks, but the main results are briefly revisited to give the reader a tutorial-like overview on the QoE analysis of M/M/1-FIFO systems. Please note that the full Kendall notation of the system is M/M/1/∞/FIFO, as the waiting room is not limited.

2.1 Concepts and Notation

The concepts and notation frequently used in the paper are summarized below in Table 1. Random variables (RV) are typically denoted by upper case letters. We first describe general concepts, before the system and QoE relevant parameters are introduced.

Table 1. Notation and variables

	Notation	Meaning
General	$\mathrm{E}[X]$	expected value of a random variable X with probability density function $x(t)$, $\mathrm{E}[X] = \int_{-\infty}^{\infty} tx(t)dt$
	$\mathrm{Var}[X]$	variance of a random variable X, $\mathrm{Var}[X] = E[X^2] - E[X]^2$
	$\mathrm{Std}[X]$	standard deviation of a random variable X, $\mathrm{Std}[X] = \sqrt{\mathrm{Var}[X]}$
	$F_X(t)$	cumulative distribution function (CDF) of the RV X, $F_X(t) = P(X \leq t)$
	$f_X(t)$	probability density function (PDF) of the RV X, $f_X(t) = \frac{d}{dt} F_x(t)$
$M/M/1 - FIFO$	λ	arrival rate of user requests in the system (1/s)
	μ	service rate of user requests (1/s) with mean service time $E[X] = 1/\mu$
	X	service time (RV) of requests (s)
	ρ	load in the system corresponding to the system utilization $\rho = \lambda/\mu$
	W	waiting time (RV) of a user in the system (s)
	R	response time (RV) of the system (s)
QoE Mapping	$f(t)$	generic mapping function $f(t) = -a \log_{10}(t + b) + H$ between waiting times and QoE, see Eq.(9), with service-dependent parameters a, b and upper QoE bound H
	Q_m	maximum possible QoE, i.e. $Q_m = \min(H, f(0))$
	t_m	largest waiting time for which QoE still reaches its maximal value, i.e. $Q_m = f(t_m) = f(t)$ for $t \leq t_m$
	t_L	minimum QoE is L for any $t \geq t_L$
	$f_C(t)$	initial delays for YouTube obtained via crowdsourcing, see Eq.(10)
	$f_L(t)$	initial delays for YouTube tested in a laboratory setting, see Eq.(11)
	$f_S(t)$	authentication in social networks, see Eq.(12)
	$f_3(t)$	wireless 3G Internet connection setup, see Eq.(13)
QoE Values	Y	QoE values (RV) obtained by mapping response times to QoE, $Y = f(R)$, thus Y is a continuous random variable
	L	lower bound of the QoE domain, i.e. $L \leq Y$
	H	upper bound of the QoE domain, i.e. $L \leq Y \leq H$, e.g. $L = 1$ and $H = 5$ for a 5-point scale
	$\mathrm{E}[Y]$	overall QoE reflecting the expected QoE Y
	$\mathcal{F}$	QoE fairness of QoE values Y

2.2 First Moments of the Response Time

In an $M/M/1$ system, let the expected service time be denoted $E[X] = 1/\mu$, where X is the random variable for the service time. The expected sojourn time in the system is then

$$\mathrm{E}(R) = \frac{1}{1-\rho}\frac{1}{\mu} = \frac{1}{\mu - \lambda},\ \lambda < \mu\ . \tag{1}$$

The expected queueing, or waiting, time $E[W]$ is then the expected sojourn time in the system minus the expected service time.

$$\mathrm{E}(W) = \mathrm{E}(R) - \mathrm{E}(X) = \frac{\lambda}{\mu(\mu - \lambda)},\ \lambda < \mu\ . \tag{2}$$

The expected waiting time given that the customer has to wait (delayed customer) $\mathrm{E}(W \mid W > 0)$ is found using the law of total expectation.

$$\mathrm{E}(W \mid W > 0) = \frac{1}{\mu - \lambda} = \mathrm{E}(R),\ \lambda < \mu\ . \tag{3}$$

2.3 Response Time Distribution

In contrast to the expected times in the system, the time distributions depend on the queueing discipline. In this paper, we consider the simplest case of FIFO queueing for three cases of waiting (queuing) time distribution, (i) waiting time for a "tagged" customer who sees q customers ahead on arrival, (ii) waiting time for customers who have to wait, and (iii) response time for all customers.

Specific (tagged) delayed customer. First, we consider the conditional waiting time distribution for customers who are delayed due to queuing. Assume that the system is in state $i = q + 1$, where $q \in N$ is the number of customers in queue immediately before a customer enters the system. This customer has to wait $q + 1$ service completions before being served. When the server is busy, the system completes customers with a constant rate μ (negatively exponentially distributed service times). Therefore, the waiting time for a customer that has to wait (delayed customer) given that there are q customers ahead in the queue is Erlang-$(q + 1)$ distributed. The cumulative distribution function (CDF) is then

$$F_{W|W>0}(t \mid q) = \sum_{j=q+1}^{\infty} \frac{(\mu t)^j}{j!} e^{-\mu t},\ q \in N,\ t > 0\ . \tag{4}$$

Conditional for customers who have to wait. The unconditional waiting time distribution for delayed customers is derived using the *law of total probability.*

$$F_{W|W>0}(t) = \sum_{q=0}^{\infty} F_{W|W>0}(t \mid q) p_{W>0_q} \tag{5}$$

where $p_{W>0_q}$ is the probability that there are q customers in the queue, i.e., $q+1$ customers in the system, immediately before a new customer enters the system given that the server is busy, i.e., given that this new customer has to wait.

Hence, in an $M/M/1/\infty/FIFO$ system the waiting time distribution for delayed customers is negatively exponentially distributed with parameter $(1-\rho)\mu = (\mu-\lambda),\ \lambda < \mu$, thus

$$F_{W|W>0}(t) = 1 - e^{-(\mu-\lambda)t}\,. \tag{6}$$

This implies that if a customer counts the number q of customers in the queue when entering the system the waiting time is Erlang-$(q+1)$ distributed with parameter μ, while if not the waiting time is negatively exponentially distributed with parameter $(\mu-\lambda)$.

Response time distribution Following a similar approach as for the waiting time distribution, if an $M/M/1$ system with FIFO queueing is in state i immediately before a customer enters the system, the response time for this customer is Erlang-$(i+1)$ distributed, and the response time, or sojourn time, for an arbitrary customer is negatively exponentially distributed with parameter $(\mu-\lambda),\ \lambda<\mu$.

Due to the PASTA[5] property [20] for a stationary system with Poisson arrivals, the probability that an arrival finds the system in state q is equal to the probability that an outside observer finds the system in state q at an arbitrary point of time. The arrival- and the time-stationary distributions are identical.

The response time distribution in an $M/M/1/\infty/FIFO$ system is negatively exponentially distributed with parameter $(\mu-\lambda),\ \lambda<\mu$. For the sake of simplicity, the CDF of the response time is also written as $R(t)$ in this paper.

$$R(t) = F_R(t) = 1 - e^{-(\mu-\lambda)t} \tag{7}$$

Figure 1 provides the CDF of the response time depending on the utilization ρ. With an increasing system load, the response times are increasing, but also the variance of the response time is increasing. Next, we will use the response time and map it to QoE values to investigate the system in a user-centric way.

3 QoE Models and QoE Fairness

In order to map the system response times to the user-perceived quality, we use existing QoE mapping functions. These are based on subjective studies [6]. The QoE mapping allows to objectively estimate QoE values in the $M/M/1$ system. The users are not differentiated, and therefore a response time r is mapped to a QoE value y for any given user. This *objective* view on the user-perceived quality allows a system provider to, e.g., do QoE management in a meaningful way. In the literature, there are several measures to quantify QoE [7]. The most commonly used QoE measure is the Mean Opinion Score (MOS), which represents

---

[5] "Poisson Arrivals See Time Averages"

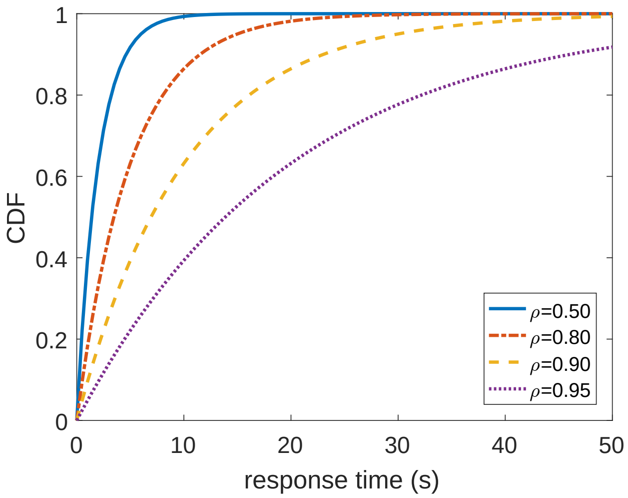

Figure 1. Cumulative distribution function (CDF) of the response times of an M/M/1-FIFO queue with different load ρ. The response time (also referred to as sojourn time) is the waiting time in the FIFO queue and the processing time of a job at the server.

the quality experienced by a hypothetical "average user". For a service provider allocating resources to users, it may however be more important to consider other measures of quality, such as the 10 % most annoyed users, which may be expressed by the 10 %-quantile. Service providers may also want to consider the percentages of users judging a service as "poor or worse" (%PoW) or "good or better" (%GoB). Those users who are experiencing lower quality than the MOS (mean) would suggest, are the ones who might be more susceptible to churn, or open help-desk tickets, etc., all of which has direct business consequences.

In this paper, we use MOS as our QoE measure, simply because there are good mappings available between response times and it in the literature [6], but in actual usage by, say, a service provider, other measures might be better suited.

For the user-centric analysis of the M/M/1 system, the response times are mapped to QoE values $Y = f(R)$ and the overall QoE as well as QoE fairness are investigated. The definition of both notions is introduced in Section 3.2.

3.1 Existing Mapping Functions between Waiting Times and QoE

The QoE of Internet applications and services is often shaped by waiting times before — or during — service consumption. Those waiting times may be a result of insufficient resources (e.g., limited transmission capacity, limited cloud computing resources), network impairments (e.g., packet loss or high latency), or simply time-consuming operations. In [6], subjective user studies were conducted to analyze the differences in the user perception of initial delays for different interactive services. In the studies, the users evaluated the QoE on a so-called 5-point absolute category rating (ACR) scale with the following meaning: 5 - excellent, 4 - good, 3 - fair, 2 - poor, 1 - bad quality. Then for each waiting time and Internet service under test, the average rating score was computed reflecting the MOS value for that test condition. Based on those subjective results, the relationship between QoE and the waiting times were derived. In [3,4], a hypothesis was formulated that the relationship between waiting time and its QoE evaluation on a linear ACR scale is logarithmic, motivated by the logarithmic form of the well-known Weber-Fechner law [19]. In [6], the mapping function is formulated as follows:

$$f^*(t) = -a \log_{10}(t+b) + H, \tag{8}$$

where the constant $H = 5$ reflects the upper bound of the ACR scale and a and b are service specific parameters obtained from subjective studies.

Please note that the mapping function maps a continuous response time $t \in \mathbb{R}^+$ to a continous MOS score $f(t) \in \mathbb{R}^+$. Thus, a response time distribution R can be mapped to a continous QoE distribution Y. Since Eq.(8) does not respect the limits of the rating scale (higher bound $H = 5$ and lower bound $L = 1$), the mapping functions needs to be refined by considering the related bounds $t_m = \max(0, 1-b)$ and $t_L = 10^{(H-L)/a} - b$.

$$f(t) = \begin{cases} L & \text{for } t \geq t_L = 10^{(H-L)/a} - b \\ -a \log_{10}(t+b) + H, & \text{for } t_L \leq t \leq t_m \\ Q_m, & \text{for } t \leq t_m = \max(0, 1-b) \end{cases} \tag{9}$$

The maximum QoE being observed is $Q_m = f^*(t_m) = -a \log_{10}(t_m + b) + H$, while the minimum QoE is L for any $t \geq t_L$. In the following, we simplify the notation and only provide the logarithmic function and the bounds.

Initial delays in YouTube video streaming. In general, HTTP streaming utilizes a video buffer to both decrease the impact of network jitter and to decrease the probability of interruptions during the video playout. For YouTube video streaming, a certain buffer level is to be reached [18] before the video starts playing. For the evaluation of the encountered initial delays, a laboratory study as well as a crowdsourcing study were conducted. The two different test methodologies are not of importance for this paper, but the small deviations in the mapping functions are of interest if they are relevant for different load scenarios in the $M/M/1$ system. As a result, the following mapping functions were found for the crowdsourcing and the laboratory setting, respectively [6].

$$f_C(t) = -0.963 \log_{10}(t + 5.381) + 5,\ Q_m = 4.2962,\ t_m = 0, t_L = 14240.40 \tag{10}$$

$$f_L(t) = -0.862 \log_{10}(t + 6.718) + 5,\ Q_m = 4.2869,\ t_m = 0, t_L = 43682.19 \tag{11}$$

Authentication in social networks. The second Internet service addresses the user authentication in social networks. In [13,12], users evaluated the perceived quality of web-based login operations using a laptop. In the subjective experiments, a remote OpenID server was run for authenticating the users as backend of the web page of the social network. The waiting times of the users were changed with a a traffic shaper. To be more precise, the shaper induced pre-determined response times for the authentication procedure when the user logged in. After the delayed login, the users were asked to rate how they experienced the login with regards to the response time resulting in the following mapping function [6].

$$f_S(t) = -2.816 \log_{10}(t + 1.378) + 5,\ Q_m = 4.6079,\ t_m = 0, t_L = 24.95 \tag{12}$$

Wireless 3G Internet connection setup. The next use case considers the perception of waiting times for wireless 3G Internet connection setup. In [15], subjective experiments were conducted where test users were sitting in front of a laptop for the 3G connection. For simulating different waiting times, a network emulator was customized in such a way, that the time span from pressing a "Connect" button to successful connection establishment was delayed for a defined time period. After the successful connection setup, the users evaluated again the QoE on a 5-point ACR scale how satisfied they were with the performance of the connection setup. The corresponding QoE mapping function is as follows.

$$f_3(t) = -1.577 \log_{10}(t + 0.742) + 5,\ Q_m = 5,\ t_m = 0.2580, t_L = 343.18 \tag{13}$$

Figure 2 illustrates the different QoE mapping functions for the different Internet services. It can be seen that the waiting time perception across different services strongly diverges. As a concrete example, let us consider an initial delay of 10 s. In case of YouTube, this leads to good QoE (3.86 crowdsourcing, 3.95 laboratory), whereas for the 3G setup users perceive this somewhere between fair and good (3.37). In case of the social network authentication, the quality is perceived as bad (2.03). Hence, considerable differences are observed for the same waiting times across services. Please note that these are all very simple examples of QoE mappings, as they are all univariate, and consider only one single aspect of quality. The same type of approach can be used to build more complex QoE models, taking more quality-influencing factors into account.

3.2 Definition of Overall QoE and QoE Fairness

Overall QoE. For a user-centric analysis of the system, the overall QoE and the QoE fairness of the system are investigated. We can define the overall QoE as the expected QoE for an arbitrary user in the system. Please note that due to the nonlinear mapping function, the overall QoE does not follow by mapping the mean response time to QoE. In the paper, the overall QoE will be derived numerically from the CDF of QoE values. Due to Jensen's inequality [9] it is known that

$$\mathrm{E}[Y] = \mathrm{E}[f(R)] \geq f(\mathrm{E}[R]) \tag{14}$$

for a convex function f which is the case for the considered mapping functions.

QoE fairness. The QoE fairness index is defined in [8] and computed over the observed QoE values Y in a system. In particular, the standard deviation $\sigma = \mathrm{Std}[Y]$ is linearly transformed. When the QoE values are given on a QoE scale with lower bound L and higher bound H, then the fairness index is

$$\mathcal{F} = 1 - 2\frac{\sigma}{H - L} \tag{15}$$

which is on the 5-point scale with $L = 1$ and $H = 5$ as used in the paper

$$\mathcal{F} = 1 - \sigma/2 \; . \tag{16}$$

The QoE fairness metric has some nice properties and has an intuitive meaning. $\mathcal{F}$ is a continuous value bounded in the interval $[0;1]$, see [8] for a formal proof. A high value of $\mathcal{F}$ if the system is QoE-fair, low values if the system is unfair. $\mathcal{F} = 1$ means perfect fairness and all users experience the same QoE. In contrast, $\mathcal{F} = 0$ is a totally unfair system. This is observed for example if one user obtains best QoE $H = 5$ and the other gets $L = 1$ in a system of two users. Please note that the QoE fairness metric is scale- and metric-independent. Thus, it does not matter if the QoE mapping function is provided on a 5-point scale or linearly transformed to any other scale, e.g., normalized values in the interval $[0;1]$. Due to its definition, the fairness index is also independent of the actual QoE level,

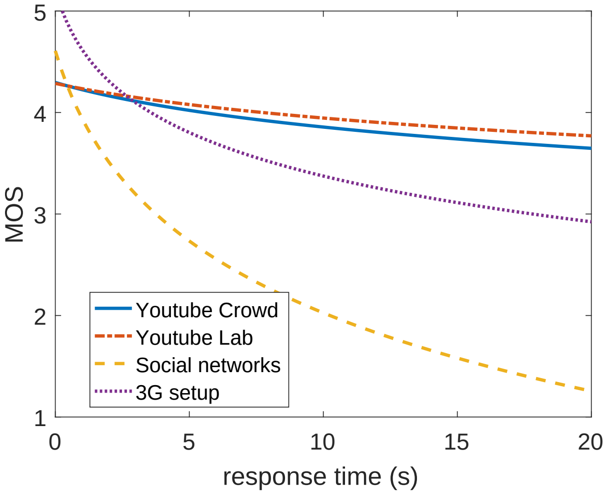

Figure 2. QoS-QoE mapping function provided in [6] for the different Internet services. The QoS parameter is the response time of the system until the service starts. YouTube considers the initial delay until the video playout starts for two different subjective studies conducted in a laboratory and via crowdsourcing. The authentication in social networks maps response times for the authentication procedure when the user logged in to MOS values. The wireless 3G Internet connection setup considers the time for a successful connection establishment and how user perceive this delay.

i.e., whether the system achieves good or bad QoE. A system can be evaluated in terms of QoE by providing the overall QoE $\mathrm{E}[Y]$ as well as the QoE fairness $\mathcal{F}$.

4 Derivation of QoE Results

For the M/M/1-FIFO system, the response time distribution $R(t)$ of a system with given λ and μ is mapped to QoE using the corresponding QoE mapping function $f(t) = y$. Hence, the QoE distribution is $Y = f(R)$ being a continuous random variable. In order to derive the CDF $F_Y(x) = P(Y \leq y)$ of the QoE values, the inverse QoE mapping function $f^{-1}(y) = t$ is required, cf. Eq.(9),

$$f^{-1}(y) = 10^{(H-y)/a} - b = e^{(H-y)/a'} - b \text{ with } a' = a/\log 10\,. \tag{17}$$

Then, the QoE distribution is as follows.

$$\begin{aligned} F_Y(y) &= P\left(Y \leq y\right) = P\left(f(R) \leq y\right) \\ &= P\left(R \leq f^{-1}(y)\right) = F_R\left(f^{-1}(y)\right) \\ &= 1 - e^{-(\mu-\lambda)\cdot\left(e^{(H-y)/a'} - b\right)} \end{aligned} \tag{18}$$

Although the analytical solution of the QoE distribution is specified, the expression for the overall QoE (first moment), as well as for the QoE fairness requiring the second moment leads to rather complex equations. Therefore, the results are derived numerically. The PDF $p(y) = \frac{dF_y}{dy}$ is numerically derived based on the complex-step derivative approximation [14].

The overall QoE is then numerically derived by taking into account the bounds of the QoE scale, see for example the PDF in Figure 3,

$$E[Y] = \int_{-\infty}^{\infty} y \cdot p(y)dy = \int_{1}^{Q_m} y \cdot p(y)dy + L \cdot P_L + Q_m \cdot P_m \tag{19}$$

with the probability for the lower bound $P_L = P\left(Y = L\right) = P\left(R \geq t_L\right)$ and the probability of the upper bound $P_m = P\left(Y = Q_m\right) = P\left(R \leq t_m\right)$. Please note that the upper bound may be a value $Q_m < 5$, see Eq.(9). In a similar way, the second moment $E[Y^2]$ is derived which allows to compute the variance $\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2$, standard deviation $\mathrm{Std}[Y] = \sqrt{\mathrm{Var}[Y]}$, and QoE fairness $\mathcal{F} = 1 - \mathrm{Std}[Y]/2$. Please note that the symbolic math toolbox from MATLAB®was used to exactly compute the (lengthy and complex) expressions for the first and second moments of the QoE values which are omitted here. Instead, only the numerical results are provided in the following section.

5 Numerical Results

In this section we briefly discuss the results of joining the performance analysis of the $M/M/1-FIFO$ system with the QoE models described in Section 3.

Previously, on Figure 1, we saw how the response time distribution for the $M/M/1$ system varies with the load ρ. We can also see, in Figure 2, how the QoE mappings approximate the quality perceived by the users, for the three Internet services, as a function of the response time.

In Figure 3, we can see the CDF and PDF of the QoE estimates (MOS values, in this case) in a case where the system is highly loaded ($\rho = 0.9$). These results were obtained by composing the QoE mappings with the response time distribution observed for in the $M/M/1 - FIFO$ system with the given load.

In Figures 4 and 5, we can see the overall QoE and QoE fairness, respectively, for the $M/M/1 - FIFO$ system, as a function of the system load. As the reader probably noticed in the previous figures, there are clear differences between services in how the users perceive the impact of response time on the quality. In particular we note that the social network login case shows the worst quality of the lot. When looking at the fairness of this service, we can see that it goes up at the end of the load scale (after ~ 0.9). This indicates that a large proportion of users is already experiencing the lowest possible quality at that stage, and hence the fairness goes up once more (remember that QoE fairness is independent of the overall QoE; it only reflects the variation in quality observed among users).

Finally, Figure 6 plots the overall QoE against QoE fairness. We can observe, once again, very different behavior between the social media login case, and the others, as well as an overall lower fairness of both the 3G setup and the social media login cases when compared to the video streaming ones (which is to be intuitively expected, as some initial delay in video streaming is common and thus expected by the users). The variation in the shape of the QoE fairness curves can also be related to the distribution of the QoE scores, as observed in the PDF plots in Figure 3, where both the social media login and 3G connection setup show a larger variability in the scores, as well as a higher skew towards the lower end of the quality scale.

6 Conclusions and Outlook

This tutorial paper introduces a framework for the user-centric analysis of queueing systems in which response times are mapped to QoE values. For the end user, those response times manifest as waiting times before service consumption. As a simple example, an M/M/1-FIFO system is investigated for the setup of different Internet services: 1) YouTube until the video playout starts, 2) authentication in social networks, 3) wireless 3G Internet connection setup. For the analysis of the system, the overall QoE as well as the QoE fairness are used. These two measures allow for example to properly dimension a system such that the users obtain a good QoE while achieving fairness in terms of QoE among users. The numerical results suggest that the interpretation of the system behavior in terms of QoE significantly differs for certain services. But it can also be seen that for the dimensioning of the service rates it is sufficient to consider the overall QoE only. A lower overall QoE reduces also the QoE fairness up to a certain point. This arises from the fact that higher system load in FIFO queues

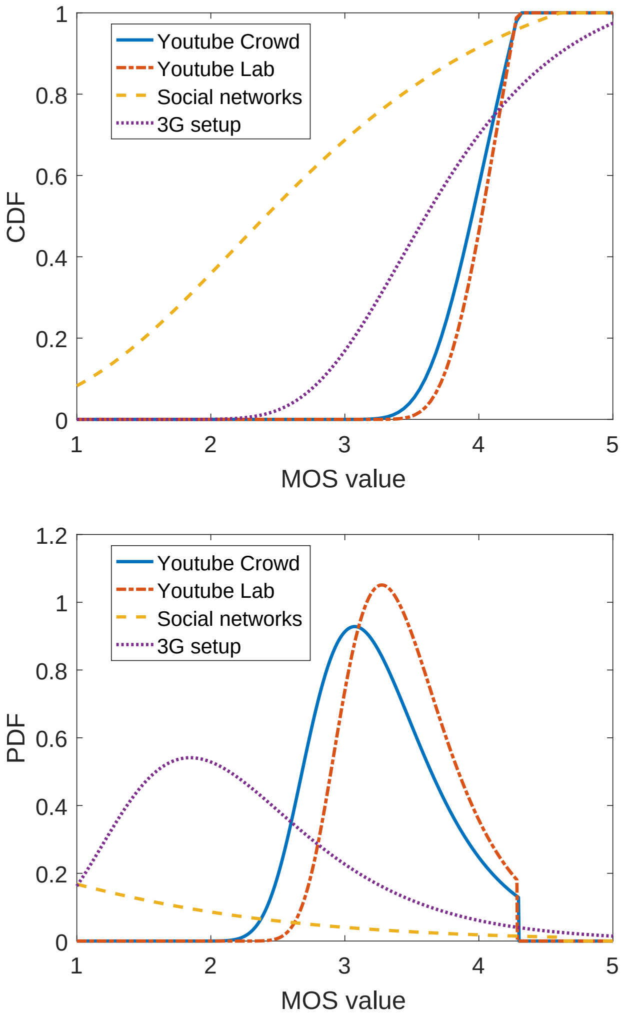

Figure 3. CDF and probability density function (PDF) of the QoE quantified as MOS values for different applications with a system load $\rho = 0.9$. The QoE mapping functions (see Figure 2) were applied to the distribution response time R observed in an M/M/1-FIFO queue with $\rho = 0.9$.

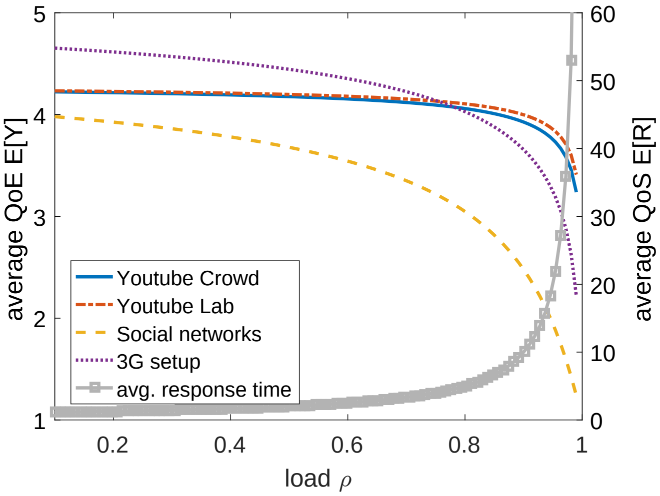

Figure 4. The overall QoE of the system is expressed as average QoE $E[Y]$ over all users who experience QoE Y. The QoE Y is a random variable which is a function of the response time R, i.e., $Y = f(R)$ with the corresponding mapping functions $f(r)$ for the different services. The average response time $E[R]$ is plotted on the right y-axis depending on the various system utilizations ρ.

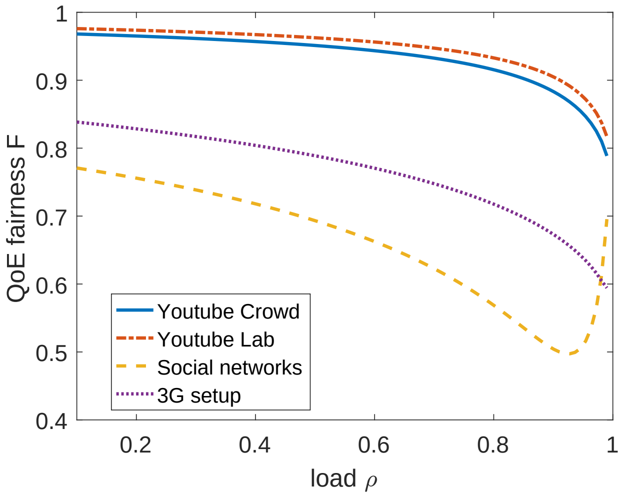

Figure 5. The QoE fairness $\mathcal{F}$ of the system is investigated for different services.

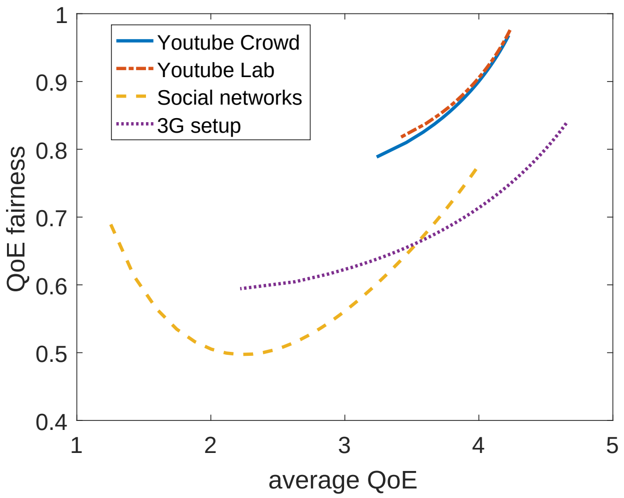

Figure 6. Scatter plot of the average QoE $E[Y]$ and the QoE fairness values $\mathcal{F}$.

also leads to higher variances in the response time. If the load in the system exceeds a certain threshold, then the overall QoE is poor or even worse. When all users are suffering, the fairness increases due to decreased variances in QoE, but the system is not working in an acceptable way for the end users.

Future work will address different scheduling strategies to evaluate them in terms of overall QoE and QoE fairness. To this end, it is also interesting to investigate more sophisticated QoE metrics like 10 %-quantiles or ratio of satisfied users which may be more appropriate for QoE dimensioning. Nevertheless, the same framework may be followed to analyze such systems with different metrics. This type of analysis can also be extended to consider other types of systems, as well as other types of services. For example, bounding the system capacity (i.e., an $M/M/1/K$ system) allows us to consider other performance aspects beyond time, such as the loss process in the network (e.g., deriving loss rates and average loss burst sizes from the system's load), which in turn allow us to consider other QoE models for e.g., real-time media applications, for which losses are a very important influencing factor. In the case of interactive media services, both delays and losses are important, and this type of approach allows us to attack this problem.

References

1. Altman, E., Barakat, C., Ramos, V.M.R.: Queueing analysis of simple FEC schemes for IP telephony. In: Proceedings of INFOCOM '01. pp. 796–804 (2001)
2. Brunnström, K., Beker, S.A., De Moor, K., Dooms, A., Egger, S., Garcia, M.N., Hossfeld, T., Jumisko-Pyykkö, S., Keimel, C., Larabi, M.C., et al.: Qualinet white paper on definitions of Quality of Experience (2013)

3. Egger, S., Hoßfeld, T., Schatz, R., Fiedler, M.: Waiting times in quality of experience for web based services. In: Fourth International Workshop on Quality of Multimedia Experience (QoMEX). pp. 86–96. IEEE (2012)
4. Egger, S., Reichl, P., Hoßfeld, T., Schatz, R.: “Time is bandwidth”? Narrowing the gap between subjective time perception and Quality of Experience. In: IEEE International Conference on Communications (ICC). pp. 1325–1330 (2012)
5. Emstad, P.J., Heegaard, P.E., Helvik, B.E., Paquereau, L.: Dependability and Performance with Discrete Event Simulations. Textbook in MSc course, "TTM4110 Dependability and Performance with Discrete Event Simulations". Norwegian University of Science and Technology (NTNU). Unpublished (2016)
6. Hoßfeld, T., Egger, S., Schatz, R., Fiedler, M., Masuch, K., Lorentzen, C.: Initial delay vs. interruptions: Between the devil and the deep blue sea. In: Fourth International Workshop on Quality of Multimedia Experience (QoMEX). pp. 1–6. IEEE (2012)
7. Hoßfeld, T., Heegaard, P.E., Varela, M., Möller, S.: QoE beyond the MOS: an in-depth look at QOE via better metrics and their relation to MOS. Quality and User Experience 1(1), 2 (2016)
8. Hoßfeld, T., Skorin-Kapov, L., Heegaard, P.E., Varela, M.: Definition of qoe fairness in shared systems. IEEE Communications Letters 21(1), 184–187 (2017)
9. Jensen, J.L.W.V.: Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta mathematica 30(1), 175–193 (1906)
10. Kleinrock, L.: Queueing systems. Volume 1: Theory. Wiley-Interscience, 1 edn. (1975)
11. Kobayashi, H., Mark, B.L., Turin, W.: Probability, random processes, and statistical analysis. Cambridge University Press (2011)
12. Lorentzen, C., Fiedler, M., Johnson, H.: On user perception of safety in online social networks. International Journal of Communication Networks and Distributed Systems 11(1), 77–91 (2013)
13. Lorentzen, C., Fiedler, M., Johnson, H., Shaikh, J., Jørstad, I.: On user perception of web login—a study on QoE in the context of security. In: Australasian Telecommunication Networks and Applications Conference (ATNAC). pp. 84–89. IEEE (2010)
14. Martins, J.R., Sturdza, P., Alonso, J.J.: The complex-step derivative approximation. ACM Transactions on Mathematical Software (TOMS) 29(3), 245–262 (2003)
15. Reichl, P., Egger, S., Schatz, R., D’Alconzo, A.: The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment. In: IEEE International Conference on Communications (ICC). pp. 1–5 (2010)
16. Rubino, G., Varela, M.: Evaluating the Utility of Media–dependent FEC in VoIP flows. In: LNCS 3266: Proceedings of the Fifth International Workshop on Quality of future Internet Services (QofIS’04). Barcelona, Spain (9 2004)
17. Skorin-Kapov, L., Varela, M.: A multi-dimensional view of QoE: the ARCU model. In: 2012 Proceedings of the 35th International Convention, MIPRO 2012, Opatija, Croatia, May 21-25, 2012. pp. 662–666 (2012)
18. Wamser, F., Casas, P., Seufert, M., Moldovan, C., Tran-Gia, P., Hossfeld, T.: Modeling the YouTube stack: From packets to Quality of Experience. Computer Networks 109, 211–224 (2016)
19. Weber, E.: Annotationes anatomicae et physiologicae [anatomical and physiological annotations]. Leipzig: CF Koehler (1851)
20. Wolff, R.W.: Poisson Arrivals See Time Averages. Operations Research 30(2), 223–231 (1982)