Mapping QoE with Resource Estimation in IoT

Mohammad Aazam and Khaled A. Harras
Carnegie Mellon University
azam@ieee.org, kharras@cs.cmu.edu

Abstract—Technologies associated directly or indirectly with Internet of Things (IoT) are growing rapidly, paving way for multimedia-based IoT services such as social IoT (SIoT), healthcare/telemedicine, and Tactile Internet. Multimedia services require a certain level of quality of service (QoS), which can be a challenge in the case of resource-constrained devices, particularly, mobile nodes. Therefore, it becomes imperative to dynamically allocate resources incorporating quality of experience (QoE). In this paper, we provide a methodology for incorporating QoE in different ways such as overall QoE and the QoE provided by the current customer. QoE is determined through net promoter score (NPS). Our mathematical model determines the ratio with which the resources need to be scaled up to meet the desired QoS. We have implemented our model using Java, and tested its impact using the CloudSim simulator. The results show the effect of the QoE ratio determined by our proposed model.

Index Terms—IoT; Multimedia; QoE; QoS; Resource Management

I. INTRODUCTION

Multimedia-based Internet of Things (IoT) applications is gaining importance as the technologies around multimedia communication are becoming cheaper and and more applications rely on more heterogeneous IoT devices. With the availability of 4G long-term evolution (LTE) communication, and soon to be available 5G, more high-quality video communication will be enabled. As mobile communication also advances, along with agile virtualization techniques (e.g. containers), coupled with futuristic paradigms such as Tactile Internet (TI), Industrial Internet of Things (IIoT) [1], cyber-physical systems (CPS) [2], and intelligent transportation systems (ITS), video communication will be an inevitable part of smart services. More than entertainment and standard communication, video or multimedia-based communication will be playing a vital role in the healthcare sector (e.g. remote surgery, remote patient monitoring) [3].

Several services will have strict quality of service (QoS) requirements, and quality of experience (QoE) will play a vital role in determining the required QoS. Examples include: remote surgery via Tactile Internet, remote patient monitoring, drone applications (e.g. delivery, surveillance, or emergency management), IIoT, and CPS. If a patient is monitored/operated remotely or a drone is distantly maneuvered, user experience will determine the amount of resources required to be allocated in order to reach user satisfaction. Hence, more adaptive and dynamic multimedia delivery methods are needed. Currently, only 17 percent of Internet video is supported by adaptive or dynamic streaming technologies [4]. Dynamic methodologies will not only achieve the desired level of QoE but will also help improve resource utilization.

Multimedia content often requires high processing power, storage, bandwidth, and better resource scheduling. Henceforth, it becomes important to manage these resources effectively to perform efficient resource management at the service provider (e.g., cloud, fog, or edge server) [5]. Especially with IoT nodes that do not typically experience reliable/consistent connectivity quality, the cost considerably increases when it comes to resource allocation and a huge amount of resources go underutilized because of the unexpected and unreliable behavior of the user [5].

In this paper, we build on our previous work [5] where we proposed a basic mathematical model for IoT resource estimation (RE), with the focus on customer historical records. Here, our focus is on QoE-based dynamic resource estimation, referred to as QoE Ratio (QoER). QoE has a direct impact on the cloud and IoT business process since the reputation of the service provider, provider’s net profit, and mitigation of resource underutilization depend on it. There are different ways to acquire QoE and one of those is through net promoter score (NPS) [6], which we incorporate in our work.

Customers provide NPS-based QoE feedback on a scale of 0-10. Where customers giving a score between 0 to 6 are known as detractors, 7-8 are passives, and 9-10 are promoters. Final NPS is determined by subtracting detractors from the promoters. In the default case when there is no feedback available, there are no promoters and detractors. The default NPS in that case is the mean of passive, which is 7.5.

We determine NPS ratio \( NPS_c \) using two parameters. One is the overall NPS \( NPS_o \) which is given by all the users who have previously used the service \( S \). The other parameter is the NPS of the particular customer \( C \) \( NPS_c \), currently requesting a service \( S \). We analyze the effect of \( NPS_c \) on the basis of which dynamic RE can be performed. Such a system can play a vital role in dynamically improving the QoS according to the changing conditions of the service. The service quality can be close-to-desired, and resource underutilization can be mitigated, resulting in profit for the service provider. We implement our model in Java and provide empirical results using the CloudSim simulator.

In rest of the paper, section II is on prior related works. Section III presents the proposed QoE Ratio model. In section IV, we present the outcomes of the proposed model. Section V concludes the paper with some insights on the future work.
II. RELATED WORK

Reiter et al. [7] discuss influence factors (IFs) in QoE. IFs can be defined by characteristics of a user, service, or context, that may influence a user’s QoE. There are three categories of IFs: human IFs (HIFs), system IFs (SIFs), and context IFs (CIFs), as shown in Fig 1. IFs may influence each other. For instance, HIFs and CIFs might determine to what extent SIFs affect QoE; a video watched on a mobile phone, or on a TV, might have different QoE. HIFs are defined by a user’s emotional, physical, and/or mental state, demographic and socio-economic back-ground. SIFs are the attributes that define the technically produced QoS. CIFs correspond to a situational property describing a user’s environment.

Hossfeld and Keimel [8] emphasize that traditional QoE evaluation has limitations, one of which is the unavailability of the number of simultaneous participants. Moreover, demography of the subjects is often not illustrative of the diversity in the general population. The evaluation setup also does not truly represent most real-life environments. One viable solution to this problem is to crowdsource QoE evaluations over the Internet via user feedback. This way, more realistic QoE evaluation can be performed. Yitong et al. [4] examine the aspects that influence the QoE of adaptive streaming, and assess it when running an end-to-end service. It has been endorsed that with respect to QoE, adaptive streaming greatly improves end-user’s subjective perception, as compared to the fixed-rate streaming. Li et al. [9] presented a QoE evaluation for IoT on the basis of linear regressions analysis. They analyze the type of IoT applications, determine the applicable QoS parameters of QoE, and collect the sample data accordingly. However, dynamically adaptive methodologies are not discussed in this work.

Floris and Atzori [10] presented mean opinion score (MOS)-based QoE evaluation for multimedia IoT. A layered IoT architecture is analyzed in their work to get relevant QoE IFs required according to relevant application scenarios. Ning and Wang discussed the potential of IoT and the amount of data it is going to generate [11]. The authors also emphasize efficient management of resources for the future Internet in which heterogeneous IoTs would be an essential part.

III. PROPOSED QoER MODEL

In this section, we present our QoER model for resource estimation. This model will be implemented where the resources are allocated for a service, such as cloud, fog, edge server. Customers make a request for a service \( S \). The service provider in a cloud-IoT scenario makes sure that the QoS requirements are met. Customer’s feedback (QoE) is considered while estimating the resources, so that the service quality can meet the expectations of the user. An NPS ratio is generated using the overall NPS of a service and the NPS provided by the current customer. Later, resources are estimated according to the determined NPS ratio. In this way, resources are always allocated according to the expectations and changing requirements of each customer. This not only improves the service quality and customer satisfaction, but also works for the marketing of the service and net profit gain for the provider. Additionally, it helps to mitigate resource underutilization [12].

During the resource estimation phase, resources are increased whenever QoS requirements are not met in the previous case of service consumption, so that better service is provided and customer’s loyalty is gained, as the resource consumption goes on. NPS ratio is calculated as follows:

\[
NPS_r = \begin{cases} 
\sum_{i=0}^{n} \frac{NPS_{d}}{\pi NPS_{oi}}, & \text{if } n = 0 \\
\sum_{i=0}^{n} \frac{\pi NPS_{oi}}{NPS_{d}}, & \text{if } \pi NPS_{oi} > NPS_{d} \\
\sum_{i=0}^{n} \sum_{k=0}^{n} \frac{\pi NPS_{oi}}{NPS_{ck}}, & \text{if } \pi NPS_{oi} \geq \pi NPS_{ck} \\
\sum_{i=0}^{n} \sum_{k=0}^{n} \frac{\pi NPS_{ck}}{\pi NPS_{oi}}, & \text{if } \pi NPS_{ck} \geq \pi NPS_{oi}
\end{cases}
\]

(1)

\( NPS_r \) is the NPS ratio determined on the basis of the overall NPS (\( NPS_{oi} \)) of the service instance \( i \) that is currently requested by a customer with a possible certain historical NPS (\( NPS_{c} \)). There are four cases: Case 1 represents the situation when the customer is new, without any previous NPS record. In this case, the default NPS (\( NPS_{d} \)) is applied which is 7.5. However, it is possible that the requested service may have a higher overall NPS (\( NPS_{oi} \)) than the default value 7.5. It shows that majority of the users have a better opinion about the service quality, hence, higher priority should be given to their opinion. This corresponds to case 2 of our model. When a customer is returning (previous NPS records exist), either of case 3 or 4 is applied. Case 3 corresponds to the situation when a customer \( c \) with \( k \) number of previous instances of service utilization (\( NPS_{ck} \)) has a higher value than the overall NPS (\( NPS_{oi} \)). In the otherwise situation, case 4 is applied. The ratio increases as the gap between \( NPS_{ck} \) and \( NPS_{oi} \) increases, so that the resources are scaled up according to the particular opinion that would eventually enhance service quality. In the case when (\( NPS_{ck} \)) and (\( NPS_{oi} \)) are the
same, ratio remains 1 because it shows that everyone has the same experience, hence, a particular customer does not need exclusively greater resources for better QoS, rather the service in general needs to be scaled up. Thenceforth, in such a case, regular, linear increase in resources will be applied.

All types of NPS are calculated through Eq. 2, in which \( NPS_{pr} \) represents the promoters and \( NPS_{dt} \) represent the detractors.

\[
NPS = NPS_{pr} - NPS_{dt} \tag{2}
\]

IV. IMPLICATIONS AND OUTCOMES

In this section, we provide details on implementation of our proposed model, it evaluation, and the results of the evaluation, along with their discussions.

A. Evaluation Setup and Implementation

We implement and evaluate our proposed model using Java on the CloudSim simulator. The evaluation setup is based on real IBM datacenters (IBM server x3550 (2 x Xeon X5675 3067 MHz, 6 cores], 16 GB) defined in the CloudSim toolkit. We generate 1000 different requests for service instances with different QoS requirements (delay, jitter, packet loss and so on). We use a Gaussian distribution (as it is more generic and continuous probability distribution) to generate different request types which represent different instances of the services. We generate the QoE feedback randomly in the form of NPS through the toss−of−a−die methodology. We accommodate different service instances requested by the customers with different historical statuses, such as new customer with no record available, existing customer with high QoE feedback record, and existing customer with low QoE feedback record. All these factors are incorporated while evaluating our model.

B. Performance Metrics

The metrics used in the evaluation are as following:

Overall NPS: Average NPS of a particular service provided by all the customers that have used the service so far.

Default NPS: This is the default, initial NPS, which is the average of the passive range (7.5).

Customer’s Historical NPS: This is the average NPS of the current customer that has requested the service.

C. Results

We now present our evaluation results on QoE ratio as part of the overall resource estimation methodology.

1) Default NPS: Fig. 2 shows the NPS ratio \( (NPS_c) \) which is based on the current customer’s average NPS \( (NPS_c) \) and the overall \( (NPS_o) \). However, as it is the default case, there is no previous NPS record for customer \( c \). Thenceforth, the default NPS \( (NPS_{dt}) \) which is 7.5 according to Eq. 2 is applied. The horizontal axis in Fig. 2 shows different \( NPS_o \) for each service, and accordingly, different \( NPS_c \). The larger the gap between \( NPS_o \) and \( NPS_c \), the higher the \( NPS_r \), which implies that additional resources with that ratio are required to improve the QoS. For example, for service 1, when \( NPS_o \) is 7.5, the \( NPS_r \) becomes 2.5. Hence, additional resources with the ratio of 2.5 will be required to improve the QoS, or otherwise the service will continue to decline in its QoE. In the case of service 5 where \( NPS_o = 9 \) (higher than the default case of 7.5), the \( NPS_r \) becomes 0.84. The reasons why this RE is less than 1 is because the customer is new, without any previous NPS history. Which means that the service provider does not know the customer’s loyalty. If the customer quits the service, the resources will be underutilized. Hence, resources have to be allocated cautiously in the beginning. It does not mean that the customer will experience a comparatively bad service, but resources would be increased as soon as the service continues to ensure there are enough resources for other reliable existing customers.

2) When \( NPS_o > NPS_c: \) In this case the customer is returning, and his/her historical record \( (NPS_c) \) exists, which is essentially case 3 of Eq. 1. To evaluate this case, let us assume that the average overall NPS \( (O\bar{NPS}_o) \) is 8. In the first instance in Fig. 3, the overall and customer specific NPSes are the same. This means that the quality most of the people have experienced for this service is the same as that of the current customer. Hence, no change in the ratio. In the second instance when the customer’s NPS is 4, the \( NPS_r \) becomes 2. Similarly, \( NPS_r \) becomes 2.67 as the gap between \( NPS_c \) and \( NPS_o \) increases. The ratio shrinks as the gap between \( NPS_c \) and \( NPS_o \) decreases in instance 4. In the same way, the resources are scaled up based on the ratio determined, which results in enhanced service quality every time a customer requests for it. This eventually plays its role in determining customer’s loyalty, resulting in improving resource utilization.

3) When \( NPS_c > NPS_o: \) When the current customer’s NPS is greater than the overall NPS (case 4 of Eq. 1), \( NPS_r \) will be affected accordingly. Opposite to the previous case, on this occasion, since \( NPS_c \) is already higher than \( NPS_o \), apparently, there is no need to enhance resources. Nevertheless, the reason why \( NPS_r \) still matters here is...
because if the current customer gets a better service, it would result in even better QoE, which would eventually enhance the overall QoE. This is imperative for marketing purposes. Fig. 4 shows case 4 of Eq. 1. As an example, \( NPS_o \) is 7 for service \( s \) and \( NPS_c \) varies according to different utilization instances. In the first instance \( NPS_c \) is 8 which is slightly higher than the overall NPS. The \( NPS_r \) becomes 1.15. This means that the difference between the general opinion by others is not much from the current customer. Therefore, there is only a slight increase in resources. But if the difference increases, it would mean that the general opinion needs to be bettered as well, in order to enhance the overall QoE. However, there is an exception to it in the case when \( NPS_c \) is 10 (which is perfect QoE). This exceptional case is represented through instance 6. As the current customer has no complaint with the QoE, there is no point in increasing the resources without any reason. As a result, the increase ratio \( (NPS_r) \) is zero. Here, it should be noted that zero means no increase at all, 1 means a linear or default increase (according to the policy of a service provider), and rest of the ratio determined through our model tell the exact factor with which the resources have to be increased.

V. CONCLUSION AND FUTURE WORK

The rising importance of IoT-based services has resulted in adoption of more sophisticated ways to provide improved services that meet rising user QoE expectations. QoS is a major concern when it comes to multimedia services, that includes services related to telemedicine, remote healthcare, and various Tactile Internet applications. Resource allocation in currently deployed systems does not take into account QoE while estimating resources. If QoE is considered, better QoS can be maintained. As such, a fair business process is upheld, reliable customer behavior is gained, resource underutilization is minimized, and profit is increased. Therefore, it is a win-win situation for all the involved parties. This is what we have focused on in this paper. We provided a methodology that takes into account the overall QoE and a specific customer’s QoE, on the basis of which a ratio is calculated. Resources are dynamically estimated on the basis of this ratio. Our work provides some insight on how dynamically QoE-based resource estimation can be done to achieve the desired service quality results. An extended part of this work is in press [13], in which we focus on Tactile Internet applications. We plan to further extend this work on one or more use-cases of Tactile applications.

ACKNOWLEDGMENT

This publication was made possible by NPRP grant # 8-1645-1-289 from the Qatar National Research Fund, a member of Qatar Foundation. The statements made herein are solely the responsibility of the authors.

REFERENCES