QoE-centric Service Delivery: A Collaborative Approach among OTTs and ISPs

Arslan Ahmad, Alessandro Floris and Luigi Atzori

DIEE, University of Cagliari, Italy

Abstract

The provisioning of the quality to end users is a major objective for the successful deployment of multimedia services over the Internet. It is more and more evident from past research and service deployments that such an objective often requires a collaboration among the different parties that are involved in the delivery of the service. This paper specifically focuses on the cooperation between the Over-The-Top (OTTs) and the Internet Service Providers (ISPs) and propose a novel service delivery approach that is purely driven by the Quality of Experience (QoE) provided to the final common users. Initially, we identify the need of the collaboration among the OTTs and the ISPs where we not only highlight some of the enterprise level motivations (revenue generation) but also the technical aspects which require collaboration. Later, we provide a reference architecture with the required modules and vertical interfaces for the interaction among the OTTs and the ISPs. Then, we provide a collaboration model where we focus on the modeling of the revenue, whose maximization drives the collaboration. The revenue is considered to be dependent on the user churn, which in turn is affected by the QoE and is modeled using the Sigmoid function. We illustrate simulation results based on our proposed collaboration approach which highlight how the proposed strategy increases the revenue generation and QoE for the OTTs and the ISPs hence providing a ground for ISP to join the loop of revenue generation between OTTs and users.

Keywords: Quality of Experience, QoE, ISP, OTT

1A preliminary version of this paper has been presented at QoMEX 2016
1. Introduction

Internet traffic has evolved over the past decade from web traffic to multimedia traffic due to the widespread use of smartphones as multimedia content generators and significant advancement in multimedia services over the Internet. Recent studies on trends in Internet traffic have predicted that more than 75% of the world’s mobile data traffic will be multimedia by 2020 [1]. Such a drastic increase in the use of multimedia services requires more resources at the Internet Service Providers (ISPs) end to assure the required level of quality to the users, although ISPs are not in the loop of revenue generation between the providers of the multimedia services (i.e., the OTTs - Over-The-Top providers) and the users. Indeed, the ISPs, as well as the OTTs, are affected by the reaction of the users to low service quality as they are more and more quality aware. Then, both entities must face the risk of user churn which may result into decrease of market share and reputation which is unavoidable in this era of strong competition in this domain.

The recent researches over the past years have revealed that quality perceived by the users not only depends on quality delivered by the network but also on application parameters and subjective factors. Indeed, the Quality of Experience (QoE) is a multidimensional concept in which several influencing factors are involved, such as: human, context, price and application aspects. Accordingly, the eco-system for QoE delivery analyzed in [2] shows that application and network providers are key players involved in QoE delivery and both contribute to the final quality level delivered to the users.

Lately, the research in the field of QoE has been conducted separately with the different prospects of the OTTs and the ISPs which gave birth to two different areas in the field of QoE: application-aware networks and network-aware applications. The network-aware applications [3] aim to adapt the delivery of multimedia contents on the best effort over the network by inducing change in the application parameters, whereas application-aware networks [4] focus on effective management of network according to application requirements. However, the drawback of the above mentioned research stream is that OTTs have no control over network for enhancing users’ QoE, whereas ISPs have no availability of application model neither encrypted content nor users’ privacy allow them to go for deep packet inspection [5]. Hence, both the ISP and the OTT cannot deliver the best QoE to their valued customers, which results into user churn as well as decrease in market shares.

On the basis of these considerations, in this paper we focus on the inves-
tigation of the impact of a QoE centered OTTs-ISP collaboration for QoE
based service delivery to end users. At the first we discuss some of the tech-
nical aspects and impacts of collaboration highlighting the need of OTT-ISP
collaboration for QoE based service delivery. Later, on the basis of the possi-
gle roles of the OTTs (QoE monitoring and application optimization) and the
ISP (QoS monitoring, revenue maximization and network-wide operations),
we propose a reference architecture which defines the interfaces and modules
required for their interactions providing a baseline for continuous exchange
of information/service between the two entities. Then we propose the QoE
centered collaboration approach which is driven by the maximization of the
revenue based on different factors, such as the user churn (which is modeled
as affected by the QoE using the Sigmoid function), pricing and marketing
actions. The collaboration is guided by ISP which maximizes the revenue
as a function of the delivered QoE with the provision of better network re-
sources on the basis of application QoE model while the OTTs perform the
context-aware QoE monitoring and provide the ISP with the information
about the class of service per user as well as about application parameters.
Finally, with simulations we highlight how the proposed collaboration ap-
proach increases the revenue generation and the QoE for both the ISP and
the OTT.

The paper is structured as follows. Section 2 discusses the state-of-the-
art related works , while Section 3 highlights the need for collaboration.
Section 4 presents the reference architecture, whereas Section 5 discusses the
proposed collaboration approach. Section 6 provides the simulations based
on our proposed approach and finally Section 7 concludes the paper and
discusses future work.

2. Past works

This section reviews the works that propose new algorithms for QoE-
centric service delivery, those that focus on OTT-ISP collaboration, and those
that address the user churn modeling.

2.1. QoE-Centric Service Delivery

The delivery of quality in accordance with end user perception is only
possible if the service delivery is QoE centered, i.e., with the inclusion of
application specific QoE models in the service delivery process. Accord-
ingly, some of the works found in literature defined QoE centered approaches.
In [6], the authors presented a QoE monitoring model based on network and application parameters in Long Term Evolution (LTE) architecture. The provided results highlighted that different applications require different level of network resources on the basis of their QoE models. Similarly, the work presented in [7] proposed QoE based scheduling algorithm for LTE networks where higher scheduling priority is given to packets of mostly used application based on QoE models. The results shown that the VoIP and video streaming required high level of network resources in order to deliver better quality. The case of wireless LAN is addressed in [8].

In [9], Varela et al. highlighted that QoE provision to end user cannot be done with the current Service Level Agreements (SLA) but rather Experience Level Agreements (ELA) would be required to deliver guaranteed QoE. Whereas in ELA the change in Service Level Objectives (SLOs) from mean time to failure or mean time to recovery in QoS parameters to minimum assurance of Mean Opinion Score (MOS) was proposed. The work also proposed agreement between OTTs and ISPs on QoE based SLOs.

Other works addressed the pricing strategies between ISPs and OTTs. In [10], the authors investigated the cases of QoS sold by the ISP to the OTT or to the users. The impact of different QoS pricing strategies were modeled analytically and analyzed with numerical results. It resulted that the ISP may sell QoS to users at a lower price than when QoS is sold to the OTT. Similarly, the studies in [11] proposed a coalition model for CDNs and ISPs based on QoS where CDNs will pay ISPs for better provision of QoS to their traffic. In [12, 13], the authors propose a pricing model based on the network architecture similar to the Paris Metro Pricing (PMP) method proposed in [14]. The PMP aims at partitioning the main network into logically separate channels where each channel has fixed fraction of network capacity and associated price. There would be no guarantees of QoS because packets are always delivered on a best-effort basis. However, the channels with higher prices are expected to be less congested than those with lower prices, resulting in provision of better quality to customers who pay more. The study in [12] demonstrates pricing for the network with two service classes for any number of competing ISPs. From their analysis, they concluded that a network with two service classes is socially desirable, but it could be blocked due to unfavorable distributional consequences, i.e., violation of network neutrality principle. Furthermore, they demonstrated that in the absence of regulation and considerable ISP market power (small), a sizable fraction of the current network users will experience a surplus loss with two service classes. In [13],
the PMP method has been integrated with QoE aspects giving birth to PARQUE (Pricing and Regulating Quality of Experience). PARQUE considers two different types of applications (web traffic and video traffic) implying higher QoS requirements for video traffic than for the web traffic. For both types of application the users QoE expectations are considered together with the user willingness to pay for the service.

From the results provided by these studies, it can be stated that providing different classes of services to the users on the basis of their willingness to pay can improve quality as well as the revenue. However, when studying network resource allocation among different applications an important factor must always be considered, i.e., the network neutrality (also called Net Neutrality or NN). Although there is no standard definition yet, Net Neutrality principle states that in order to preserve the openness of the Internet, the end users should have equal access to all the content on the Internet, and the ISP should be prohibited from discriminating/blocking the content from any of the application providers [15]. For such principle, the network should deliver traffic in a best effort manner, but lower levels of neutrality violation can be accepted as intrinsic prioritization, load management and blocking of illegal content [15]. In [16], the authors discussed the Net Neutrality with social, economical and technical prospects where authors classify Net Neutrality as a threat to future innovation and technology which may eliminate ISPs incentives to invest in the network.

2.2. OTT-ISP Collaboration: Technology Oriented Aspects

Although the collaboration among OTTs and ISPs is catching the eyes of researchers working in QoE-oriented service management, still only few works have really addressed this aspect in the literature. The collaboration between networks and applications in the future Internet is addressed in [17], where the importance of the collaboration between network providers and applications is highlighted by discussing a scenario in which applications give more information about their needs and network usage so that ISPs can allocate network resources more efficiently or even open their network so that applications can dynamically invoke some network services. Two existing collaboration techniques are discussed: the ALTO (Application-Layer Traffic Optimization) [18] and the CINA (Collaboration Interface between Network and Application) [19]. The ALTO initiative allows P2P networks and ISPs to cooperate in order to optimize traffic being generated by P2P applications and transported over the ISPs infrastructure. However, the application-ISP
interaction in ALTO only concerns network information provided by ISPs and processed by applications, i.e., the ISP is blindfolded to the services which their customers subscribe to. These limitations are addressed by the CINA interface, which not only allows applications to retrieve information about the network, but also offers the possibility to instantiate network services such as multicast service, caching nodes, and high capacity nodes. Nonetheless, these works are specific for P2P applications and the collaboration between network and application is limited. Furthermore, business aspects are not investigated.

2.3. User Churn

According to the study conducted in [20], quality and pricing are considered as major causes for a user to become churner. Nowadays, the users’ satisfaction related to a particular service plays an important role in the growth of market share of any company dealing with multimedia services and it has high cross correlation in the prediction of users’ churn as well. However, to the best of authors’ knowledge, no works can be found in literature regarding users’ churn model in terms of quality perceived by the user. In fact, most of works propose utility functions which model the QoE on the basis of network and application parameters. For example, in [21] the Sigmoid function is used to model user satisfaction as a function of QoS parameters, such as delay and error rate, for Internet Protocol Television (IPTV). In [22], the IQX hypothesis is presented, i.e., a generic exponential relationship between user-perceived QoE and network-caused QoS. This relationship has been proved to be valid for some case studies, such as: voice quality as a function of loss and jitter; cancellation rates of web surfer as a function of access link bandwidth. Indeed, these and other related works provide a QoE measure in function of specific QoS and application parameters, i.e., they can be useful for the monitoring of end-to-end system parameters. However, what is missing in the state-of-the-art is a model which is able to estimate the influence of the QoE in causing customer churn for telecommunications services.

3. Why is OTT-ISP collaboration needed?

It is well known that the QoE for any service over the Internet not only depends on the network parameters but also on the application parameters [23]. Accordingly, as the OTTs applications are being delivered over the
ISPs best-effort Internet without considering the resource requirement of the
application, the degradation may lead to serious user churn. However, gen-
erally, the ISP is the entity which suffers more from the user churn because
the average user thinks that the poor QoE perceived is mainly due to low
network resources and then changes ISP. Additionally, the user may decide
to move from one operator to another. For this reason, OTTs are usually not
willing to collaborate with ISPs as well as because they do not want to share
any precious information about their application and users. Nonetheless, the
OTTs may accept to collaborate if this collaboration allows them to increase
their revenue, i.e., the network services provided by the ISP allow the OTT’s
users to perceive a better QoE so that the number of users of that OTT
provided through that ISP increases together with OTT and ISP’s revenue.
Therefore, the collaboration between ISPs and OTTs must require a com-
mon ground of motivation that we identify as the revenue. Especially during
the last years users are more quality demanding and fulfillment of the quality
expectations may lead to the reduction in user churn which in turn increases
the number of the customers, resulting in higher revenue for both the service
providers. Hence, we propose a collaboration approach driven by the max-
imization of the revenue based on different factors such as the user churn
(modeled as a function of the QoE), pricing and marketing actions. In the
following subsections we further discuss about the technical issues that the
collaboration is addressing.

3.1. Application-aware Traffic Engineering vs Encryption

The delivery of users’ perceived quality is a big issue nowadays consid-
ering that different multimedia applications have dissimilar requirements [24].
The proposed quality management approaches at the hands of the ISPs, such
as DiffServ [25] and IntServ [26], have their own limitation over best-effort In-
ternet [27]. Though some past works highlighted to be application-oriented,
such as in [6, 7], today, the OTT services are being encrypted with the con-
cern of the user privacy issues. This is the case for example of YouTube that
has been turned from HTTP to HTTPS, where the videos are now being
transmitted in the encrypted sessions [28]. The traffic encryption is lead-
ing to a major challenge for application-aware QoE-based Internet service
delivery as ISPs may not be able to either perform the Deep Packet Inspec-
tion (DPI) and packet marking in order to apply the core traffic engineering
concepts such as packet prioritization, traffic shaping, admission control etc.
for the multimedia traffic management [28]. Moreover, shortest path routing
concept [29] cannot be applied to delay sensitive traffic. Furthermore, as the
network resource requirements vary in accordance with QoE model of the
application, only an OTT may know the best application-aware QoE model
according to the users’ context of use.

3.2. Different Roles in QoE based Optimization and Control

As also highlighted in [30, 31], the optimization of the application rate si-
multaneously with the packet prioritization and error concealment techniques
not only can save the network resources but can also increase significantly
the QoE of the multimedia streaming services. But an important issue arises
in case a single player has to perform the control and optimization as all the
levels are not in the same hands but rather distributed, i.e., the optimization
of the application parameters is in the hands of OTT only while the ISP
has control over the network resource usage. Therefore, the collaboration is
required, which may not only results in saving the number of resources but
will also lower down the user churn.

4. Reference architecture for Collaboration

The reference scenario is composed of an ISP which provides network
infrastructures and services, and different OTTs that provide over-the-top
applications. The major aspect that links the OTTs with the ISP is the QoE
delivered to the final users, which can be selected as the core component
for building collaboration strategies towards service delivery. As a matter of
fact, the OTT is aware of users’ expectations and the level of quality they
are experiencing, thanks to the control of the software at the application
level and a close relationship with the user. Indeed, through the application
software it can monitor application parameters (such as buffer occupancy
and playout delay in video streaming applications) and context parameters
(such as the type of device and the position of the user), and can even ask
the user to fill surveys about quality satisfaction. However, it cannot have
any control on the network. On the other hand, the ISP is more focused on
QoS and controls network resources provided to all of its users; however, not
always better ISP provided QoS has a positive effect on QoE.

Therefore, since the OTT is the entity which is more QoE-oriented, a
collaboration between the OTTs and the ISP can help the ISP to implement
a QoE-aware network management for the provisioning of adequate QoE to
the end-users. Fig. 1 sketches the reference architecture of the collaboration
scheme we focus on. We provide a high-level architecture which defines a set of functional requirements that must be provided by OTTs and ISP for making possible the collaboration approach. Since it is a functional architecture, we do not provide any specification about how to implement the functional blocks nor recommendations are given about the network interfaces to be used for information exchange. We assume that multiple OTTs decide to collaborate with a single ISP. The OTTs monitor the QoE of their users using QoE models which are specific for the application they are providing to their users. This is the role of the QoE monitoring block, which measures the QoE as a function of application parameters and context parameters (extracted from user profile information) such as user location, user’s device, user’s expectations, etc. The QoE measurements are then conveyed to the ISP through a dedicated interface, together with the information about the class of service of the users. In fact, a dedicated communication channel is established between each OTT and the ISP, to allow for the transmission of information between OTTs and ISP. Such a channel is interconnected at both the ISP and OTT ends with a functional block implementing Authentication, Authorization and Accounting (AAA) functions for a secure information exchange.

On the other hand, the QoS monitoring block of the ISP monitors the QoS of the network through which all the OTT applications are provided to the end-users. QoS and QoE measurements are then received and collected by the Network management block, which runs a QoE-aware network management algorithm which aims at controlling the QoE by looking at its impact on the user churn and then on the revenue. Specifically, on the basis of a model of the revenue, the best combination of price and QoE level is computed using (5). Then, relevant requirements are taken for providing network resources to quality-demanding users for maximizing the revenue. However, these decisions must take care of not putting other users at a disadvantage neither discriminating other OTTs (for assuring Net Neutrality).

5. Collaboration model

We assume that the multiple OTTs services are passing through an ISP network and they agree upon the collaboration on the basis of the roles and interfaces defined in the reference architecture proposed in Section 4. The reason for consideration a single ISP is to simplify the treatment. However, the proposed solution can be extended to multiple ISPs scenario without
any issue of scalability. Nonetheless, the proposed collaboration requires a common ground among ISP and OTTs in the form of key models related to: revenue generation, QoE, user churn, pricing and marketing. The way these models are used in our proposal is discussed in the following subsections in the following order: the pricing model, the revenue model for the collaboration, the user churn modeling and the revenue maximization approach.

5.1. Pricing Modeling

An important point of the collaboration between the ISP and the OTTs is the definition of the pricing model, i.e., the economic rules which define how much the users should pay for accessing the combined ISP-OTT services. As analyzed in [32], network congestion brought to the birth of Smart Data Pricing (SDP), i.e., a suite of pricing and policy practices that have been proposed by operators as access pricing options instead of the traditional flat-rate model. One application of SDP is the pricing for end-user QoE, i.e., pricing strategies for matching the operator’s cost of delivering bits at
the consumer’s QoE needs for different application types at the price the customers are willing to spend. SDP approaches are mainly classified into static and dynamic models depending on whether the prices are changed in real-time or on a longer timescale. In this paper, we rely on a static model: the Paris Metro Pricing (PMP) concept for Internet pricing proposed in [14]. Although the PMP model charges different prices for different network channels, there are no guarantees of QoS for the users which pay more. However, it is expected that the channels with higher prices would be less congested than those with lower prices, resulting in higher quality provided to customers paying more.

For our pricing model, we propose an enhanced version of the PMP model which assures a minimum guaranteed quality to the users depending on the money they pay. Accordingly, we assume that the services provided by the ISP-OTTs collaborations are being offered in $J$ different levels of quality (with $j = 1, 2, 3, \ldots, J$ indexing the different levels) and different prices, and that the higher the price the better is the expected (and provided) quality. The users are assigned to a specific class in function of their willingness to pay $W^n$, where $n$ indexes the user. For simplicity, as in [13] we consider normalized $W^n$ and normalized prices so that $W^n \in [0, 1]$ and $P_{i,j} \in [0, 1]$, where $P_{i,j}$ is the price to be paid to subscribe to the $j$-th class of service of application $i$. As a general example, a user $n$ will subscribe to the service class $j$ if $P_{i,j} \leq W^n < P_{i,j+1}$.

Till here, this is a QoS-based pricing model which aims at providing higher system performance to users paying more, i.e., different Service Level Agreements (SLAs) are defined between the providers and the users as a function of their willingness to pay. However, SLAs are difficult to be understood by the users and are not directly related to their perceived quality. Therefore, inspired by the concept proposed by Varela et. al in [9], we define the quality provided by each class of service in terms of Experience Level Agreements (ELA). An ELA is defined as a special type of SLA designed to establish a common understanding of the quality levels that the customer will experience through the use of the service, in terms that are clearly understandable to the customer and to which he or she can relate. Therefore, we decided to represent the quality provided by each class of service with a star rating (from 1 to 5 stars), where the stars have the same meaning of the rating values defined by the Mean Opinion Score (MOS), i.e., 1 star means “Bad quality”, 2 stars mean “Poor quality”, 3 stars mean “Fair quality”, 4 stars mean “Good quality”, and finally 5 stars mean “Excellent quality”. However, any other
representation method can be used as an alternative.

At this point, a question arises: how can quality levels be defined? Specifically, how can the collaborating ISP-OTT providers decide which QoS and application parameters provide the users with a certain quality level? This is an important point as we know that the QoE depends on many different factors, ranging from objective QoS and application parameters to more subjective factors such as the context in which an application is used (used device, environment, time of the day, social factors, etc.) and human factors (user’s expectation and experience, user’s sensitivity, etc.). In this paper, we assume to use existing (and future) QoE models depending on the considered applications, e.g., VoIP and video streaming, which investigate how the QoE perceived by the users varies in function of network and application impairments. Some use cases are addressed in the simulation section.

5.2. Revenue Modeling

Recall that our collaboration is driven by the maximization of the revenue for both the service providers, for which we need to define an appropriate model. According with the price model proposed in the previous section, here we provide a model for the revenue computation. The OTT-ISP revenue clearly evolves over the time due to several factors, such as the price and the QoE. We then consider the revenue as a discrete-time process where \( t_x \) \( (x = 0, 1, 2, \ldots) \) indexes the time instants at which the revenue is computed and corrections to the system are introduced. Furthermore, we define the time window \( T = t_{x+1} - t_x \) as the period of time during which the prices of the classes of service and the number of users belonging to each class are static.

The combined revenue for the \( i \)-th OTT and the ISP (we are not separating the revenue) can be computed as follows

\[
R^x_i = \sum_{j=1}^{J} N^x_{i,j} \cdot P_{i,j} \tag{1}
\]

where \( i = 1, 2, 3, \ldots, I \) indexes the OTTs collaborating with the ISP and \( N^x_{i,j} \) is the total number of users belonging to the \( j \)-th class calculated at time \( t_x \) for the \( i \)-th OTT service. \( P_{i,j} \) is the price to be paid for subscribing to the \( j \)-th class of service of the \( i \)-th OTT application. Accordingly, the total revenue generated by the collaboration between the ISP and all the OTTs
can be calculated as

\[ R^x = \sum_{i=1}^{I} R_i^x = \sum_{i=1}^{I} \sum_{j=1}^{J} N_{i,j}^x \cdot P_{i,j} \]  \hspace{1cm} (2)

The evolution of \( N_{i,j}^x \) over the time depends on the churn effect, i.e., the process of users leaving the service. Hence, we consider that the users being represented by \( N_{i,j}^x \) are the active users of both the \( i \)-th OTT and ISP, i.e., the user continuing the services in \( j \)-th class. This number then evolves over time due to the churn and due to the activation of new contracts, as follows

\[ N_{i,j}^{x} = N_{i,j}^{x-1} \cdot U_{i,j} + \zeta_{i,j} \]  \hspace{1cm} (3)

where \( U_{i,j} \) is the user churn function that is defined in the Section 5.3 and \( \zeta_{i,j} \) is the number of users joining the \( j \)-th class of collaborative service of \( i \)-th OTT through advertisement. Indeed, studies conducted in [33] emphasis that mostly the companies gain their customers by effective marketing/advertisement campaigns, which is something considered in our modeling but not controlled by our strategy. Specifically, the study in [34] emphasized that the Poisson distribution can be utilized to predict the increase in the market share in telecommunication. Hence, we consider \( \zeta_{i,j} \) as a stochastic process which follows a Poisson distribution depending upon marketing strategies, socio-economic factors and product discounts.

5.3. Churn Modeling

The user satisfaction to a service plays an important role in the reputation of any service provider in the market. Lowering the level of user satisfaction may result into high level of user churn, i.e., reduction of the number \( N_{i,j}^x \) of active users. Notwithstanding the importance of this phenomenon, only limited works exist about the study of the impact of QoE on the user churn. One major obstacle is that to predict/model user churn in terms of QoE requires data over long periods of observation from both OTT and ISP. Still, to go ahead with our analysis, we consider that there is a high cross-correlation between user satisfaction and user churn and we build a user churn function based on the Sigmoid function [21]. Indeed, it is one of the mostly used activation function in Multi-layered Perceptron Neural Networks in the field of artificial intelligence to model human perception into machine [35, 36]. We consider the user churn function as upward criterion function, i.e., the function increases with the increase in QoE, which means that more users will
be continuing the service if higher QoE is provided, and vice versa. The user
churn function can be defined mathematically in terms of QoE as follows

\[ U_{i,j}(QoE_{i,j}) = \frac{1}{1 + e^{-z(QoE_{i,j} - QoE_{m_{i,j}})}} \]  

(4)

where \( QoE_{i,j} \) is the quality delivered to the \( j \)-th class of service of the \( i \)-th
OTT service, whereas \( QoE_{m_{i,j}} \) is the quality level at which half of the paying
users leave the service in the class \( j \) (i.e., \( U_{i,j}(QoE_{i,j}) = 0.5 \)). Moreover,
the sensitivity of the users with regard to the price paid is represented by
\( z \). In fact, users who pay more expect to receive a better quality than those
who pay less, and the users keeping the service for the former class of service
must be lower than that of the latter, for the same value of QoE perceived.
Hence, the higher the price paid the smaller the \( z \), i.e., higher the sensitivity
of the user with the quality. Fig. 2 shows an example of the user churn
function for different values of \( QoE_{m_{i,j}} \) and \( z \). The user churn function ranges
in the interval \([0, 1]\) where 1 means that the 100% of the users are keeping
the service. The QoE is measured as for the MOS in the interval \([1, 5]\) where
1 means minimum quality and 5 maximum quality. In the example shown in
Fig. 2 there are two different groups of curves: the continuous curves refer
to the lower class of service whose users have lower QoE expectations and for this reason although the perceived QoE is 2.5 half of the users will be keeping the service. On the other hand, dotted curves refer to the higher class of service, whose users are paying more and therefore have higher QoE expectations. In fact, in this case half of the users will be keeping the service for a MOS at least of 4, which means that half of users for being satisfied and keeping the service expect a more than good QoE. The different values of $z$ identifies the different sensitivity of the users and depends on the price paid to be subscribed to that class of service.

It is important to note that we defined the user churn function following the recommendations in [37]: 1) the user churn function follows the characteristics of the users’ QoE; 2) the user churn function does not change drastically with small changes in the QoE. Moreover, mathematically the proposed user churn function is valid in accordance with the law of diminishing marginal utility, which implies three conditions:

1. Concavity of $U_{i,j}(QoE_{i,j})$, $\forall QoE_{i,j} \geq QoE_{i,j}^m$
2. Convexity of $U_{i,j}(QoE_{i,j})$, $\forall QoE_{i,j} \leq QoE_{i,j}^m$
3. $U'(QoE_{i,j}) \geq 0$

We also want to stress that the proposed user churn function can be better calibrated when data about user behavior is available.

5.4. Revenue Maximization

With the complete modeling of the revenue we can now achieve the target of its maximization with a coordinated control of OTTs and ISP. Specifically, they target the maximization of an average revenue computed during a reference period $t_{X_1} - t_{X_2}$ as follows

$$\bar{R}^* = \max_{QoE_{i,j}, P_{i,j}} \left[ \sum_{x=X_1}^{X_2} \left( \frac{\sum_{i=1}^t (\sum_{j=1}^J (N_{i,j}^{-1} \cdot U_{i,j} + \zeta_{i,j}) \cdot P_{i,j})}{t_{X_2} - t_{X_1}} \right) \right]$$

where the influence of $QoE_{i,j}$ is taken into account by the user churn function $U_{i,j}$. Even if not explicitly highlighted here, the different combinations of the $QoE_{i,j}$ need to be considered under the available network resources. Then, the collaboration between the OTT and the ISP is fundamental for delivering adequate QoE to the users. The OTT is able to know QoE expectations of the user and to measure the QoE delivered, while the ISP supports the OTT by providing the needed network services. The output of this maximization
are the $QoE_{i,j}^*$ levels and prices $P_{i,j}^*$ for the different service classes. This maximization will be done at the ISP side in order to assure better quality to the end users in accordance with the reference architecture provided in Section 4, whereas through the interface defined between the OTTs and the ISP the maximization results can be shared with the respective OTTs involved in collaboration.

Herein, we want to highlight that how the OTT and the ISP decide to divide the revenue is out of the scope of the paper. However, a joint venture approach can be considered for the revenue sharing where an involved enterprise can get the share of revenue proportional to the amount of investment made by that enterprise over the total amount of investment for the service delivery. For an ISP, the investment may occur in the form of maintenance, up-gradation and operations of the access/core network infrastructure while an OTT can make investment in specialized data centers, multimedia streaming servers, content delivery networks and data-center networks. Additionally, it is important to note that this approach can be implemented without violating the Net Neutrality principle in several ways: better QoS could be provided by hosting OTT content in ISP nodes and using under-utilized network areas without affecting the other OTTs’ traffic; traffic could be prioritized if it does not affect the final QoE of the applications of which the traffic flows could be delayed.

6. Simulation

The objective of the conducted simulations is an analysis of the potential of the proposed collaboration approach. Specifically, we consider an ISP and 2 OTTs and we investigate their revenue generation for two different approaches: No Collaboration (NC) and Joint Venture (JV). The former does not consider a collaboration between the OTTs and the ISP so that the OTTs deliver their contents through the best effort service over the ISP network. The JV is the collaboration approach described in Sections 4 and 5, i.e., the OTTs collaborate with the ISP with the objective of maximizing the revenue. Without loss of generality, we focused on two specific OTT services: video streaming and VoIP. In Section 6.1 we discuss the QoE models used to evaluate the QoE perceived by the end-users whereas in Section 6.2 we present the simulation settings and results.
6.1. QoE models

As a use case, we consider two different OTT applications: video streaming and VoIP. The reason for the selection of the aforementioned applications is in accordance with the studies conducted in [6, 7] that considered video streaming and VoIP as the most sensitive multimedia applications with reference to network resources usage. For the evaluation of the QoE we based on the model proposed in [38] for the video streaming application and on the E-Model for the VoIP application [39].

The model proposed in [38] is a parametric packet-layer model for monitoring the video quality of IPTV services, which measures the QoE provided by HD (1440x1080) videos encoded with the H.264 codec at different bitrates and corrupted by packet loss:

\[
QoE_{HD} = 1 + \left( v_1 - \frac{v_1}{1 + \left( \frac{BR}{v_2} \right)^{v_3}} \right) \exp \left( - \frac{PLR}{v_4} \right) \tag{6}
\]

where \( v_1 = 3.8, v_2 = 4.9, v_3 = 3.6 \) and \( v_4 = 3.5 \) are the coefficients of the model while \( BR \) and \( PLR \) are the source coding rate of the video and the packet loss rate of the network, respectively. We consider this model for evaluating the QoE provided by HD videos because it considers both application (source coding rate) and network (PLR) parameters and because the cross-correlation factor computed between the proposed model and subjective QoE results is greater than 0.9 with 99% confidence interval.

In [40], the authors extended the model in [38] considering also the movement of the video content, the MPEG-2 codec and different video resolutions. The model is as follows:

\[
V_q = 1 + 4K \left( 1 - \frac{1}{1 + \left( \frac{b \cdot BR}{v_5} \right)^{v_6}} \right) \tag{7}
\]

\[
K = 1 + k_1 \exp(-k_2 \cdot b \cdot BR) \tag{8}
\]

where, for videos with medium content movement, encoded with the H.264 codec at SD resolution, \( v_5 = 0.67, v_6 = 1.4, b = 1, k_1 = 1.36 \) and \( k_2 = 1.93 \). Also the model in eq. (7) can take into account the effect of the PLR if multiplied for the exponential factor of eq. (6). Therefore, we consider the
model in eq. (9) for evaluating the QoE provided by SD videos:

\[
QoE_{video}^{SD} = 1 + 4K \left( 1 - \frac{1}{1 + \left( \frac{BR}{v_5} \right)^{v_6}} \right) \exp \left( -\frac{PLR}{v_4} \right)
\]  

(9)

Both the models in eq. (6) and eq. (9) measure the QoE with values ranging from 1 (Bad quality) to 5 (Excellent quality) as the MOS.

The E-Model is a planning parametric model defined by the ITU for VoIP applications, which measures the voice quality in terms of the R-factor, i.e., a quality index ranging from 0 to 100, where 100 is the best quality. The R-factor is defined in terms of several parameters as follows

\[
R = 100 - I_s - I_d - I_{ef} + A
\]  

(10)

where \( I_s \) is the signal-to-noise impairment, \( I_d \) is the impairment associated to the mouth-to-ear delay of the path, \( I_{ef} \) is the equipment impairment associated with the losses within the codecs and \( A \) is the advantage factor which allows for compensation of impairment factors when the user benefits from other types of access to the user. The study in [41] presented an adapted version of the E-Model (see eq. (11)), which emphasizes the effect of sources of quality degradation observed over data networks, namely one-way delay, packet loss ratio, and coding scheme. The adapted model is

\[
R = 94.2 - I_d(d) - I_e(CODEC, PLR) + A
\]  

(11)

where \( I_d \) and \( I_e \) capture the quality degradation caused by delay and equipment impairment factors, respectively. \( d \) is the mean one-way delay of played voice packets during an assessment interval, \( PLR \) is the packet loss ratio, and \( CODEC \) is the used speech encoding scheme.

The quality degradation caused by one-way delay when echoes are perfectly removed are calculated as

\[
I_d(d) = 0.024 \cdot d + 0.11 \cdot (d - 177.3) \cdot H(d - 177.3)
\]  

(12)

where

\[
H(x) = \begin{cases} 
  1, & x < 0 \\
  0, & x \geq 0
\end{cases}
\]  

(13)
On the other hand, the quality degradation caused by equipment impairment factors are calculated as

\[ I_e(CODEC, PLR) = a_1 + a_2 \cdot \ln(1 + a_3 \cdot PLR) \]  

(14)

where \(a_1, a_2\) and \(a_3\) are coefficients obtained through a logarithmic regression analysis depending on the used speech codec. For example, for the G.729a codec \(a_1 = 11, a_2 = 40\) and \(a_3 = 10\) whereas for the G.711 codec \(a_1 = 0, a_2 = 30\) and \(a_3 = 15\).

With regard to the advantage factor, the default value of \(A\) in case of conventional wirebound communication system is \(A = 0\). The maximum values of \(A\) are provided in [39] for different scenarios. For example, \(A_{MAX} = 5\) in case of mobility by cellular networks in a building and \(A_{MAX} = 10\) in case of mobility in a geographical area or moving in a vehicle.

Furthermore, in [41] is also provided an equation for converting the R-factor with values between 1 and 5 as the MOS:

\[
QoE_{VoIP} = \begin{cases} 
1, & R < 0 \\
4.5, & R > 100 \\
1 + 0.035R + 7 \cdot 10^{-6}R(R - 60)(100 - R), & 0 < R < 100
\end{cases}
\]  

(15)

We consider the model in eq. (15) for evaluating the QoE provided by the VoIP application, where \(R\) is computed with eq. (11).

6.2. Simulation results

The simulation scenario considers two OTT applications which are delivered to their users through a network owned by an ISP. For simplicity we assume that the users are stationary and located in the same area, where the Internet access is provided by the ISP. For both the approaches (NC and JV) and for both the applications (video streaming and VoIP), the users can choose between two different plans: standard plan (service class 1) at price \(P_{i,1}\) and premium plan (service class 2) at price \(P_{i,2}\), with \(P_{i,1} < P_{i,2}\). The subscript \(i\) identifies the OTT application. We consider normalized prices so that \(P_{i,1}, P_{i,2} \in [0, 1]\). Each user subscribes to one of the two proposed plans on the basis of his/her willingness to pay \(W^n_{i,j}\), where \(n\) indexes the user. We assume that the user is at least a standard user, then \(W^n_{i,j} \in [P_{i,1}, 1]\).

As a consequence, if \(P_{i,1} \leq W^n_{i,j} < P_{i,2}\) the user is a standard user, while if \(W^n_{i,j} \geq P_{i,2}\) the user is a premium user. Therefore, for the application \(i\), there
will be $N_{i,1}$ users subscribed to the standard plan and $N_{i,2}$ users subscribed to the premium plan, while the total number of users $N_i$ will be $N_i = N_{i,1} + N_{i,2}$.

On the application side, with regard to the video streaming application, standard users can watch videos only at Standard Quality (SD), i.e., with a resolution of 720x480 pixels, whereas premium users can watch videos at HD quality, i.e., with a resolution of 1440x900 pixels. On the other hand, both standard and premium VoIP users have access to standard VoIP services (calls, phone conferencing, etc.) whereas only premium users can have access to extra services such as recording functions, voicemail, etc. We selected the G.729 codec for VoIP simulations because it provides good performance and requires a low bandwidth (31.2 kbps [42]).

On the network side, there is a difference between the NC and JV approaches. In fact, while for the NC approach the applications are delivered on the best effort network, for the JV approach the ISP provides different network resources to the standard and premium users of the applications. Specifically, for the video streaming application, a minimum bandwidth of 2 Mbps and 5 Mbps is guaranteed to standard and premium users, respectively. In fact, generally a HD video is encoded at a bitrate ranging from 1.5 Mbps to 4 Mbps whereas a SD video is encoded at a bitrate ranging from 500 kbps to 2 Mbps [43]. Furthermore, a PLR lower than 0.3% is guaranteed to premium users whereas for standard users the maximum PLR will be 1.5%. These PLR values are selected on the basis of the study in [38] where the influence of the PLR on the QoE for video streaming has been investigated. With regard to the VoIP application, on the basis of the study in [41], a one-way delay lower than 100 ms and a PLR lower than 1% are guaranteed to premium users whereas for standard users the maximum one-way delay and PLR will be 350 ms and 5%, respectively. As discussed in Section 5.1, we express the quality of the service classes in terms of ELA and therefore we assume that with these network and application parameters the JV approach can provide at least a quality of 3 (Fair quality) to standard users and of 4 (Good quality) to premium users.

For the NC approach, the total bandwidth is divided in equal parts to each user with no guarantee of minimum bandwidth provided. Furthermore, for PLR and delay we consider the same maximum values selected for the standard users of the JV approach.

We conducted simulations with the Matlab software setting a starting number of users $N_{VoIP} = N_{Video} = 100$ and considering a total bandwidth of 500 Mbps. Since we based on the PMP pricing model, $P_{i,1}$ ranges from
0.1 to 0.5 while $P_{i,2}$ ranges from 0.6 to 1.0 with a step of 0.1 [44]. We consider $P_{VoIP,1} = P_{Video,1}$ and $P_{VoIP,2} = P_{Video,2}$. For each combination of the prices $P_{i,1}$ and $P_{i,2}$, we randomly assign a willingness to pay (uniform distribution between 0 and 1) to each user and on the basis of this value the user is assigned to the standard or premium classes of the VoIP and video streaming applications. We want to highlight that for simplicity we considered the willingness to pay uniformly distributed between 0 and 1. This way, the higher the price for joining the class of service the lower the number of users joining that class. As an example, in Fig. 3 we show the starting number of users in function of the prices $P_{i,1}$ and $P_{i,2}$. For example, within the population of 100 users, fewer users will join the premium class and more user will join the standard class as $P_{i,2}$ approaches 1. However, the willingness to pay distribution only influences the starting number of users joining the classes of service while the number of users keeping or leaving the service in next months depend on the user churn model based on the user’s QoE. Therefore, we expect that using different willingness to pay distributions will bring to the same revenue results in the long period.

Once the starting number of users for each class of service are assigned, we compute the starting video and VoIP revenue with eq. (2). We want to investigate the revenue of the 2 OTTs for the following 24 months by using eq. (5) for the revenue maximization and the QoE models discussed.
In the previous section for the QoE evaluation. With regard to $U_{i,j}$, we set $QoE_{i,1}^m = 2.5$ for the standard service and $QoE_{i,2}^m = 4$ for the premium service, since premium users have greater QoE expectations than standard users. In Fig. 2, we show the user churn function for different values of $z$ and $QoE_{i,j}^m$. With regard to $\zeta_{x,i,j}$, we computed it as a random number from the Poisson distribution with the mean equal to the 5% of users belonging to the $j$-th class and application $i$ at time $(x-1)$. The time range $x$ in this case is a month. Within each month, we compute 100 QoE measurements and we use the resulting average QoE for maximizing the revenue in eq. (5).

Fig. 4 shows the revenue obtained with the two approaches by the two OTTs in the first 2 years as a function of the prices $P_{i,1}$ and $P_{i,2}$. We did not provide the graphs of all the combinations of prices to save space, but the graphs provided allow to understand how the revenue evolves with the time for major scenarios. The most evident result is that for each prices combination the revenue obtained with the JV approach is always greater than that obtained with the NC approach for both the video and VoIP applications. This is mainly due to the fact that with the JV approach the OTTs collaborating with the ISP are able to satisfy the QoE expectations of both the standard and premium users. Specifically, the premium users are those who contribute to the revenue difference between the two approaches. In fact, standard users are less QoE demanding and they are the main contributors to the revenue generation in the case of the NC approach. Indeed, from the graphs, it is evident that when the price for the premium service is accessible to many users ($P_{i,2} = 0.6$ and $P_{i,2} = 0.8$), the NC approach fails to satisfy premium users, resulting in a great revenue drop, which is balanced over the time only thanks to the revenue provided by the standard users. When the price for the premium service reaches the highest value (i.e., $P_{i,2} = 1.0$), the standard users are prevalent with respect to the premium users and the difference between the two approaches is less evident although the JV approach provides quite higher revenue for both the applications.

From Fig. 4, it can also be noticed another interesting result concerning the $z$ parameter, which represents the sensitivity of the user to the price, as shown in Fig. 2. In fact, with the increasing of $P_{i,1}$, the standard users become more QoE demanding and are more likely to leave the service if the QoE provided is not adequate. Indeed, for $P_{i,1} = 0.4$ and $P_{i,1} = 0.5$ the revenue is not more increasing over the time as for the lower values of $P_{i,1}$, but is decreasing because not all the users are satisfied by the quality of the perceived service.
Figure 4: Revenue obtained with the two approaches by the two OTTs in the first 2 years as a function of the prices $P_{i,1}$ and $P_{i,2}$.

Fig. 5 shows the QoE provided by the video and VoIP applications for the NC and JV approaches. “ST” and “PR” stand for standard and premium service, respectively. With regard to the VoIP application for the NC approach there is no distinction between standard and premium services be-
cause the considered QoE model is a function of the only network parameters and in the case of NC approach the ISP does not guarantee any network parameter to premium users. Then, the same QoE is provided to standard and premium users. The difference between standard and premium users in this case are the extra application features which cannot be evaluated with current QoE models.

The QoE values are the average QoE computed over all the simulation cycles, and the error bars show the minimum and maximum QoE values provided. It is evident that the JV approach is able to provide a great and stable QoE to the premium users of both the applications, which results in a significant revenue generation as discussed before. On the other hand, the NC approach fails in this objective, providing to premium users a QoE even lower than that provided to standard users. With regard to the standard users, the two approaches provide comparable QoE for both the applications.

Concluding, with regard to the JV approach, the best trade-off is obtained for $P_{i,1} = 0.3$ and $P_{i,2} = 0.6$, with a quite constant revenue with an average of 40 and 35 for video and VoIP application, respectively. On the other hand, for the NC approach the most convenient prices are $P_{i,1} = 0.3$ and $P_{i,2} = 1.0$, with an increasing revenue with an average of 35 and 45 for the video and VoIP application, respectively. However, with these prices and considering

Figure 5: QoE provided by the video and VoIP applications for the NC and JV approaches.
the low QoE provided to premium users, it does not make any sense to offer
two service classes to the users but it would be better to restrict to the only
standard service.

7. Conclusion and Future Work

In this paper, we proposed a novel approach for collaborative QoE man-
agement between OTTs and ISPs. Differently from QoS based collaboration
models found in literature as ALTO and CINA, the proposed collaboration
approach is completely QoE centered based on maximization of revenue.
Moreover, the proposed model also takes into account important factors such
as the user churn, pricing and marketing, making it novel. Also, with the
consideration of the different QoE models for different applications we inves-
tigated the flexibility and adaptability of our collaboration model which has
proven to be robust, reliable and adaptable with respect to any change in
QoE model. Furthermore, with simulations we highlighted that if ISPs and
OTTs adopt the proposed collaboration approach they will not only increase
the revenue but will also provide better QoE to their users with relatively
lower prices.

Though the QoE based service delivery requires the collaboration among
OTTs and ISPs, the research in this domain is suffering from key challenges.
One of these is that no inter-operable interface exists to date which con-
tributes towards scalability of the approach. Hence, it will not only require
standardized interfaces among OTTs and ISPs to exchange QoE based infor-
mation but it will also require standardized interfaces among ISPs like peering
connections or exchange points to share QoE related information. Therefore,
the future research should focus on the provision of QoE-centric interfaces
between OTTs-ISPs to enable them QoE based service delivery. The Soft-
ware Defined Networks (SDN) and Network Function Virtualization (NFV)
can provide an opportunity in this regard because of their programmability
and flexibility. However, scalability and security remains as an open issue
in collaboration even if centralized SDN controller will be used for the QoE
management. The computational complexity for the QoE measurements may
appear to be another issue contributing to the scalability of the collaborative
approach and complexity may increase with the increase in the number of
OTT applications and customers. Moreover, there will be the requirement of
storing data related to QoE, user churn and revenue generation which may
increase the cost of network planning and operations as well.
Additionally, although big efforts have been conducted in QoE modeling, most of these are developed for providing an estimation of the perceived quality for very short periods of time. This aspect raises a practical issue when applying the resulting models to the considered scenario where the QoE affecting the churn should provide the level of experience quality resulting from longer periods of service consumption. Hence, the development of a robust and reliable QoE models valid for longer periods of time is essential for QoE based service delivery. Notwithstanding the importance of user churn, no model has been proposed which can correlate user churn with QoE which is important at the enterprise level. Nevertheless, the creation of user churn prediction model will be requiring the real customer data and analysis of that data over the significant periods of time. Moreover, the Network Neutrality and user privacy is also another future challenge for collaborative QoE based service delivery. All these challenges need to be taken into account in the future research, so that QoE based service delivery can be possible.

8. Acknowledgment

The work on this paper was partially funded from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 643072, Network QoE-Net (http://qoenet-itn.eu).

References


