QoE-aware OTT-ISP Collaboration in Service Management: Architecture and Approaches*

ALESSANDRO FLORIS†, ARSLAN AHMAD, and LUIGI ATZORI, University of Cagliari, Italy

It is a matter of fact that the Quality of Experience (QoE) has become one of the key factors that determine whether a new multimedia service will be successfully accepted by the final users. Accordingly, several QoE models have been developed with the aim of capturing the perception of the user by considering as many influencing factors as possible. However, when it comes to adopting these models in the management of the services and networks, it frequently happens that no single provider has an access to all the tools to either measure all the influencing factors related parameters or control over the delivered quality. In particular, it often happens to the Over The Top (OTT) and Internet Service Provider (ISP), which act with complementary roles in the service delivery over the Internet. On the basis of this consideration, in this paper we first highlight the importance of a possible OTT-ISP collaboration for a joint service management in terms of technical and economic aspects. Then, we propose a general reference architecture for a possible collaboration and information exchange among them. Finally, we define three different approaches, namely: joint-venture, customer lifetime value-based, and QoE-fairness-based. The first aims to maximize the revenue by providing better QoE to customers paying more. The second aims to maximize the profit by providing better QoE to Most Profitable Customers (MPCs). The third aims to maximize QoE fairness among all customers. Finally, we conduct simulations to compare the three approaches in terms of QoE provided to the users, profit generated for the providers and QoE fairness.

Additional Key Words and Phrases: Over The Top service providers, OTT, Internet Service Providers, ISP, Quality of Experience, QoE, QoE Management, OTT-ISP collaboration

ACM Reference Format:
Alessandro Floris, Arslan Ahmad, and Luigi Atzori. 2018. QoE-aware OTT-ISP Collaboration in Service Management: Architecture and Approaches. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, 1 (February 2018), 23 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

The current continuous increase of Internet traffic is mostly due to the developments in the multimedia industry conducted by Over The Top (OTT) service providers (e.g., YouTube, Netflix, Skype, Facebook) as well as to the widespread use of powerful multimedia mobile handheld devices, such as smartphones and tablets. Indeed, with the recent findings, the Internet traffic is predicted to be made for 78% by video traffic by 2021 [10]. Such a drastic increase of multimedia traffic has required Internet Service Providers (ISPs) to upgrade their network infrastructures and invest in

*A preliminary version of this paper has been presented at the QoE Management Workshop of the IFIP/IEEE International Symposium on Integrated Network Management 2017
†The corresponding author

Authors' address: Alessandro Floris, alessandro.floris@diee.unica-it; Arslan Ahmad, arslan.ahmad@diee.unica.it; Luigi Atzori, l.atzori@ieee.org, University of Cagliari, DIEE, Cagliari, 09123, Italy.

1551-6857/2018/2-ART $15.00
https://doi.org/0000001.0000001

the deployment of new technologies (e.g., optical fibers, 4G/5G mobile networks) to provide the means to handle such a huge amount of traffic.

However, network oversizing is not the right way to deal with the problem of traffic delivery because various are the types of services delivered over the Internet with different bandwidth requirements: from messaging services, which require few tenths of bytes per second, to virtual reality streaming that may require even hundreds of Mbps. Also, it would be difficult for the ISP to transform the investment in upgraded networks to revenue because the ISP is on a disadvantaged economic position with respect to OTTs, for four main reasons: i) ISPs are not in the loop of revenue generation between OTTs and users; ii) OTTs are in the market of traditional business of ISPs, e.g., voice and messaging services are also provided by OTTs such as WhatsApp and Skype; iii) although on average ISP's profit per customer is greater than OTT's profit per customer, OTTs are able to reach a much larger amount of potential customers over the globe; iv) ISPs are the most affected entities by the user churn because most of the time the user perceived quality is degraded due to bottlenecks in the ISPs' networks and the users decide to switch their Internet connection to another ISP. Consequently, this leads to decrease in market share and reputation as well as lower revenue for an ISP.

For these reasons, ISPs have to rethink current technical and business approaches, which are linked by the Quality of Experience (QoE), i.e., the quality as perceived by the user [23]. Indeed, if users perceive a high QoE, they tend not to leave the service and to continue to contribute to the ISP's profit; but the provision of high QoE for each type of service requires the ISP to adopt novel QoE-aware traffic management approaches that must be aware of the type of traffic to be managed [8, 31, 36–38]. Additionally, OTTs need to provide high QoE to their customers to increase their business: this has allowed for the growth of third entities (e.g., Akamai, Limelight, L3) that provide services, such as Content Delivery Networks (CDNs), Web Acceleration, Caching, which help to provide multimedia services with better quality. However, emerging multimedia services include high quality videos, cloud applications and highly interactive applications, which require increasingly network resources. For these reasons, OTTs and CDNs are trying to place their platform for service provision inside the ISPs' networks to be as close as possible to the end users.

But acting independently it is difficult for ISPs and OTTs to deliver adequate QoE to their customers, because QoE is a function of some parameters on the hands of the ISPs (QoS) as well as some other parameters on the hands of OTTs (application parameters, human and context factors). Therefore, a collaboration among the OTTs and ISPs seems to be the natural solution which can increase their profits with better provision of quality to their customers. However, to date research in this field is mainly divided into two separate paths: on the one hand OTTs make use of network-aware application management approaches, which aim to adapt application parameters on the basis of the monitored network status [32]; on the other hand, ISPs make use of application-aware network management approaches, which aim to manage network resources on the basis of the services to be delivered through the network [24, 31]. Some approaches towards collaboration between networks and applications for improving QoE are discussed in [30].

In this paper, we define a general architecture to allow for collaboration among ISPs and OTTs. We then propose three different collaboration approaches among OTTs and ISPs and we compare them in terms of QoE provided to the users, profit generated for the providers and QoE fairness. Indeed, one of the ultimate goals in future multimedia networks is to provide a fair user-centric quality, so that the user QoE is properly handled for all users in a network. The reason is that it may happen that selfish applications maximize its own QoE, potentially worsening QoE levels of users of different applications. According to the general fairness metric defined in [21], which satisfies QoE-relevant properties, a system is defined absolutely QoE fair when all users receive the same QoE value.

The main contributions of this work and its importance in comparison with related works are:

- We discuss the technical and economic reasons which push ISPs and OTTs to collaborate as well as their different views on QoE management.
- We propose a general reference architecture for collaboration and information exchange between ISPs and OTTs. The peering agreements and CDNs are considered for content delivery whereas the concept of InterCloud has been considered to allow the providers to exchange information with common infrastructures (cloud) and APIs.
- We propose three different collaboration approaches, namely joint-venture, CLV-based (Customer Lifetime Value - based) and QoE-fairness-based. The first, proposed in [6], aims to maximize the revenue by providing better QoE to customers paying more. The second aims to maximize the profit by providing better QoE to Most Profitable Customers (MPCs). The third aims to maximize QoE and QoE fairness among all customers. We provide the pseudo code algorithm for each of these collaboration approaches.
- We conduct simulations to compare the three proposed approaches in terms of QoE provided to the users, profit generated for the providers and QoE fairness.

The paper is organized as follows. In Section 2, we present the background regarding current collaboration agreements for service delivery and economic aspects of QoE management as well as the practical obstacles to achieving the collaboration. Section 3 highlights the different perspectives of ISP and OTT with regard to QoE management. Section 4 presents the proposed reference architecture for collaboration and information exchange among ISP and OTTs. Section 5 presents the three proposed collaboration approaches based on revenue/profit maximization (joint-venture and CLV-based approach) and QoE fairness maximization (QoE-fairness-based approach). In Section 6 the results of the conducted simulations are discussed whereas Section 7 concludes the paper.

2 BACKGROUND

In this Section, we present the background regarding current collaboration agreements for service delivery and economic aspects of QoE management which are used in the following treatment. Finally, we discuss the practical obstacles to achieving the collaboration.

2.1 Collaboration agreements for service delivery

Although a few collaboration approaches between network and application providers have been proposed in the literature for specific services such as ALTO (Application-Layer Traffic Optimization) [3] and CINA (Collaboration Interface between Network and Application) [2] for P2P networks, and SAND (Server And Network Assisted DASH) [13] for Dynamic Adaptive Streaming over HTTP (DASH). General collaboration approaches that can be used to jointly manage different types of multimedia services are still missing in the literature.

In the typical Internet service delivery chain, different key players are involved including ISPs, OTTs, CDNs and Internet eXchange Points (IXPs). The role of each player is different and the keys for better service delivery are identified as interconnection (or peering) agreements and CDNs. Peering agreements regard the interconnection between two separate networks, which allow exchanging traffic between the users of each network, whereas CDNs replicate contents on replicas of the original servers, referred to as *surrogate servers* (SSs), to lower content retrieval latency [27]. On the basis of peering policies, peering agreements can be classified as Settlement-Free-Interconnection (SFI) and Paid-peering. The former is the mostly used peering strategy and assumes that the traffic is balanced between the parties, while the latter has a cost that is typically lower than transit cost but may be considered as a violation of Network Neutrality as it may be interpreted as prioritized service. Technically, the peering can be categorized as Direct or Private

Table 1. Some of the most relevant OTTs that have direct peering with various ISPs in the U.S. [34].

OTT	AT&T	Comcast	Verizon	CenturyLink	Sprint
Google	X	X	X	X	X
Amazon	X	X		X	
Facebook	X	X	X		X
Microsoft	X	X	X	X	
Netflix		X	X	X	
Apple	X		X		

Network Interconnection (PNI) and Public Peering Interconnection (PPI) through the IXPs. Peering agreements established by some of the most popular OTT services, i.e., Google, Skype (Microsoft) and Netflix, are discussed in the following. YouTube video streaming is mostly delivered by Google[1] CDNs, which are connected to the Google core data centers. Google provides both PNIs and IXPs based peering to ISPs based on SFI peering policy with two different Google Autonomous Systems (ASs), which provide a complete set of Google services (common peering option) and a subset of Google's most popular content (available at a small number of locations). Similarly, for Skype, Microsoft[2] is providing SFI based PNIs and IXPs peering connections: the peer should declare all the routing paths and Microsoft carries traffic optimization for the traffic generated by Skype on the basis of network latency. In the case of Netflix, there are several SFI based open peering PNIs and IXPs connections for the ISPs; however, the ISP ready for the peering agreement needs to join the Netflix Open Connect Program, whose purpose is not only to provide better connection between Netflix's CDN and ISP but also to move the most popular contents closer to the end users inside the host's network with Netflix's SSs called Embedded Open Connect Appliances (OCAs)[3]. In Table 1, some of the most relevant OTTs that have direct peering with various ISPs in the U.S. are summarized [34].

Currently, most of OTT services use different CDNs for the decentralization of the contents to lower the content retrieval latency. Indeed, CDNs replicate contents on different sites and possibly on different independent ISP networks so that Web requests can be re-directed to one of the SSs according to some selection rule, which are typically based on the usage trends which are known to the OTT [27]. The SSs provide mutual benefits for the two providers: by decreasing the transit requests, the ISP decreases transit costs as well as the network overload in the backbone network; furthermore, it allows the OTT to deliver the contents with less latency than just a simple peering. Most of the collaborations based on SSs are SFI. An important case is the one of Google that is providing Google Global Cache (GGC) edge nodes to ISPs with some Google's applications (e.g., YouTube, Google Search, Google Maps), which are capable of treating from 60% to 80% of the traffic generated by Google's applications. Another is the case of Microsoft that is providing SSs to ISPs for Skype services, whereas Netflix is providing the collaborating ISPs with OCAs SSs.

2.2 Economic aspects of QoE management

It is a matter of fact that the proper management of QoE brings to direct economic advantages; indeed, if a customer does not receive adequate QoE, it is more likely to become a churner. A survey conducted by Nokia showed that around 82% of customer defections are due to frustration over the service and inability of the operator to deal with this effectively [25]. Moreover, frustrated

[1]https://peering.google.com/
[2]https://www.microsoft.com/Peering
[3]https://openconnect.netflix.com

customers can influence other customers, resulting in negative publicity for the provider. It is worth mentioning that most of the customers do not complain before defecting but they simply leave once they become unsatisfied. Therefore, it stands critical for providers investigate user's expectations and constantly monitor delivered quality to prevent the user churn. From another survey conducted by Accenture, 20% of respondents reported that they would immediately leave a company because of poor service experience [5]. But most importantly, it resulted that the number of customers who leave because of poor customer experience is significantly higher than the number of those who leave a business because they found a lower price elsewhere: 68% versus 53%. Also, a survey of 2,000 Apple iPhone users in the UK and the U.S. has revealed that the 39% of respondents would pay more for a better mobile video-streaming service [12]. Indeed, over half (59%) of subscribers in both countries will abandon streaming a mobile video if they have to wait longer than 15 s, whereas nearly a fifth (19%) will abandon a video after only a five-second wait. Therefore, service quality has a great importance in telecommunications and multimedia services and should be strongly considered to contrast the user churn and increase the revenue.

Usually, the most used parameter for revenue measurement is the Average Revenue Per User (ARPU), which is calculated as the ratio of the total revenue to the total number of subscribers. However, this metric has several issues: i) it does not explicitly represent the impact of a single user on the total revenue; ii) it makes no consideration of the service profitability and does not consider the cost of managing the customer service within the network; iii) it was defined as a voice-based index while with the advent of smartphones and tablets mobile data services are becoming prevalent and new metrics are needed to consider data and voice services separately. Indeed, it is important for the providers to understand the *value* of each customer in terms of profit, i.e., the revenue (price paid by the customer for the services received) minus the costs (costs for providing the services and managing the customer). Indeed, for the providers, some customers have more value than others and the identification of these customers allows the operator to drive retention actions to the most profitable customers (MPCs). We then consider the Customer Lifetime Value (CLV) as the metric for customer selection because it is defined as *"the total value of direct contributions and indirect contributions to overhead and profit of an individual customer during the entire customer life cycle, from the start of the relationship to its projected ending"* [16].

The CLV can be computed as a function of several parameters such as the total customer revenue, the company profit margin and the number of subscriber's loyal years. However, several modified versions of CLV formula are provided in the literature. For example, in [14] the CLV is a function of all the transactions a customer i will make for q products the company is selling. Then, for the horizon h from the period t, the CLV is defined as (note that for presentation convenience h and q are not omitted in the left part of the equation)

$$CLV_{i,t} = \sum_{k=1}^{h} \sum_{j=1}^{q} \frac{CF_{i,j,t+k}}{(1+r)^k} = \sum_{k=1}^{h} \sum_{j=1}^{q} \frac{\pi_j x_{i,j,t+k}}{(1+r)^k}, \tag{1}$$

where $CF_{i,j,t}$ is the net cash flow generated by a product j sold to a customer i during period t as a function of the product usage $x_{i,j,t}$. π_j and r are respectively the marginal profit by a unit of product usage and the discount rate. This is a marketing equation for a general product but it may be used, for example, for computing the CLV of the customers of call services, for which the profit depends on the service usage. However, this equation cannot be applied in the case of flat rates because of the pricing structure that charges a single fixed fee for a service, regardless of usage. Since flat rates are common in telecommunication and multimedia services, we provide a

CLV equation for this pricing strategy as follows

$$CLV_{i,t} = \sum_{k=1}^{h} \sum_{l=1}^{L} \frac{\pi_l y^s_{i,l,t+k}}{(1+r)^k}, \tag{2}$$

where l is the level at which a service s is provided and the marginal profit π_l depends on the level l. $y^s_{i,l,t} = 1$ if the customer i uses the service s at the level l. Otherwise $y^s_{i,l,t} = 0$. Indeed, generally, more than one fixed fee can be chosen by the customers, which depends on the level of usage of the service they need. If this level is overtaken, additional fees are applied.

One of the major objectives of ISP and OTT is to maximize their profit and to do this we focus on minimizing the churners. Specifically, for the CLV-based collaboration approach, the aim is to avoid that the MPCs become churners, because the cost of winning a new customer is much higher than retaining an existing one [29].

2.3 Practical obstacles to achieving the collaboration

It is important to highlight some practical obstacles to achieving the collaboration between OTTs and ISPs, such as Network Neutrality (NN) and privacy concerns. The respect of the NN should be considered while implementing solutions for network resource allocation among different applications to preserve the openness of the Internet. Indeed, based on the NN principle, 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 [28]. Therefore, the network should deliver traffic in the best effort manner; however, lower levels of neutrality violation may be accepted as intrinsic prioritization, load management and blocking of illegal content. Accordingly, network management approaches which result in treating differently the flows to/from collaborating OTTs should be carefully treated in this respect.

With regard to the privacy concerns related to disclosing information, the users should be involved and informed by the providers to request their consensus in the sharing of some specific information related to her activity and preferences, even if that would mean to receive higher service quality. As an incentive, the users may be rewarded with discounts or service promotions. For example, in [7], a survey is conducted to investigate users' preferences and privacy concerns to allow the provider to install monitoring probes in the user terminals. It resulted that the 28% of the users do not want to install any probe under any condition, whereas 72% agreed on the probe installation in exchange for better quality (25%), discounts on the services (10%), extra bandwidth/data usage (11%) and a combination of these 3 benefits (26%).

3 PROBLEM DEFINITION

The objective of QoE management is the optimization of end-user QoE, while making efficient (current and future) use of network resources and maintaining a satisfied customer base (provider perspective) [8]. QoE management relies on an accurate monitoring and prediction of the QoE perceived by the users, so that QoE-aware optimization and control techniques can be applied to the available resources to maximize the user's perceived QoE. In this section, we discuss ISP and OTT different perspectives in this regard, which are mainly influenced by their roles.

The ISP is the owner of the *last-mile* network through which the end-users are allowed to connect to the Internet and then subscribe to OTT services. Nowadays, users are more quality demanding so that the ISP has the difficult role in managing the enormous amount of traffic generated by OTT services (especially multimedia content) to provide adequate QoE to the users. Application level throughput and download time may be considered the most important network parameters for the delivery of current multimedia applications. The former is the actual transmission rate of the information when the end user utilizes an application or accesses a web content: it mainly depends on latency and packet losses, which typically increase with the distance between the end user and the server. Caching platforms, such as transparent caching and CDN, can improve the throughput by placing contents near to the end users. The download time represents the actual time of information download (web pages, contents), which mainly depends on the protocol efficiency (rules for information exchange between end user and server) and it is also related to the previous parameter. To decrease the download time, techniques such as protocol optimization and web acceleration platforms may be used. Though, besides best-effort services, to improve the QoE the ISP typically implements QoS-based traffic management techniques, such as the DiffServ model, which is mainly based on packet prioritization and flow management [1]. Although these actions are effective for congested network paths, they are not enough to guarantee higher throughput as well as lower delay/packet loss/download time. To this, the transfer of the multimedia servers closer to the users seems to be one of the best solutions.

The ISP has a network-centered view of the quality provided and predicts the QoE on the basis of QoS-to-QoE models, which are a function of monitored Key Parameter Indicators (KPIs), i.e.,

$$QoE^{ISP} = f_1(KPI_1, KPI_2, \cdots, KPI_a, \cdots, KPI_A), \tag{3}$$

which means that the QoE predicted by the ISP is a function of A network KPIs. Examples of QoE-based network management approaches are provided in [24, 31]. Though, the QoE predicted with Eq. (3) does not represent the actual quality perceived by the users because the QoE depends on QoS factors as well as on application, human and context factors, which are not considered in QoS-to-QoE models [40]. Indeed, the ISP does not have direct access to users' devices nor application information can be acquired by inspecting network traffic due to the data encryption applied by the OTTs. Even if the Deep Packet Inspection (DPI) may work for some of the OTT services, the human and context influence factors would be still missing and more complex context-aware QoE models for specific applications cannot be integrated into the service delivery [20]. Therefore, the provision of high QoE for each type of service requires the ISP to adopt novel QoE-aware traffic management approaches (e.g., CDN, transparent Internet caching, Web acceleration platform) that must be aware of the type of traffic to be managed as well as of user- and context-related information.

On the other hand, the OTT has access to some types of user-related information having its application running on the users' devices, from which the OTT may even monitor the most important network KPIs (e.g., throughput, delay and packet loss) and application KQIs (e.g., stalling events and playout buffer status) as well as understanding about the context in which applications are used, such as the user position (acquired from the GPS sensor), the type of device, the type of Internet connection. User profiles are also known by the OTT because each user to subscribe to the OTT service must have its own account, which contains information such as service's price and preferences and in some cases also user characteristics, which can be useful to implement a better prediction of the user's perceived QoE. Although all this information is in the hands of the OTT, the OTT may be limited in providing a good QoE to the users because of issues which may occur at the client device, at the ISP's network or in the CDN. However, even if the OTT has no control on the ISP's network, different optimization techniques may be implemented by the OTT to optimize user's QoE, such as adaptation of application parameters, CDN selection strategies, and resource management at the cloud/server side. Then, we express the QoE predicted by the OTT as a function of B Key Quality Indicators (KQIs), which in turn for the most part depend on the network QoS and on CDN strategies.

$$QoE^{OTT} = f_2(KQI_1, KQI_2, \cdots, KQI_b, \cdots, KQI_B / QoS, CDN), \tag{4}$$

Examples of QoE-based application management approaches are rate adaption algorithms for HTTP Adaptive Streaming (HAS) [32] and CDN technologies for content-centric networking [27].

From Eqs. (3) and (4), it is clear that not only the OTT and ISP have different perspectives in terms of QoE measurements but they also have different roles in the management and optimization of the delivered quality. However, optimal end-to-end QoE-based service delivery can only be possible if OTT and ISP collaborate and agree upon a common QoE-based management approach and sharing of the QoE-oriented information. Indeed, by acting independently it is difficult for ISPs and OTTs to deliver adequate QoE to their customers. Therefore, a collaboration between the two providers seems to be the natural solution to provide a better QoE to their customers. Since the OTT is more QoE-oriented, its role in the collaboration may be the monitoring of the QoE perceived by the users. By sharing this information with the ISP, this may lead to QoE-aware network management. Also, by sharing the information regarding the type of multimedia services delivered over the network, and their respective QoS requirements, the ISP does not need to employ DPI techniques but can directly provide specific network resources as a function of the service to be delivered. Then, by following a joint service management paradigm, both the entities may overtake their limitations in QoE management: the ISP can manage its network in a QoE fashion and the OTT can rely on the ISP if better network resources are needed for service delivery.

Many challenges arise from the implementation of a joint QoE management. Specifically, as suggested in [8], four basic questions should be answered: (1) what to control? (2) where to control? (3) when to control? and (4) how to control? In Section 4, we provide a reference architecture for collaboration between OTTs and ISP where we try to address these questions as well as to further questions which arise from the considered collaboration such as: (5) what information is shared between OTT and ISP? (6) when and how frequently this information is shared? (7) how this information is shared? Moreover, we consider a further question which goes beyond the joint collaborative management, i.e., what is the objective of QoE management for the providers? Indeed, there are different objectives that may drive the management of the QoE of the users such as provide better QoE to users paying more (for revenue/profit maximization) or provide the same level of QoE to each user (QoE fairness maximization). These strategies are crucial for ISP and OTTs because the providers cannot neglect the economic aspect. But, as discussed in Section 2, nowadays the economic aspects are strictly related to the delivered quality and user churn which requires a compromise between profit maximization and fair service delivery. After presenting a reference architecture for collaboration, in the following we propose three collaboration approaches which focus on the different possible QoE-related objectives.

4 REFERENCE ARCHITECTURE FOR COLLABORATION

In this Section, we present our proposed reference architecture for collaboration among ISP and OTTs, which is illustrated in Figure 1. This figure highlights the modules, information and network elements involved among ISP and OTTs to drive the joint QoE management of network and applications. Although just one OTT is shown in Figure 1, the collaboration approach can consider multiple OTTs involved in the collaboration approach with the ISP that provides the transport services to the end-users. As discussed in Section 3, OTT and ISP have different perspectives with regard to QoE management, and with the proposed collaboration we identify the role of the OTT as the QoE monitoring whereas the ISP monitors the QoS and manages network resources.

To accomplish the OTT's main task of QoE monitoring, specific QoE measurements techniques for the delivered services are needed, which would allow for a real-time prediction of delivered QoE, as a function of network KPIs and application KQIs, typically based on parametric models, deep learning and No-Reference methods [36–39]. However, to provide a more reliable estimation of the QoE, such QoE-measurement solutions may need accurate information about the network

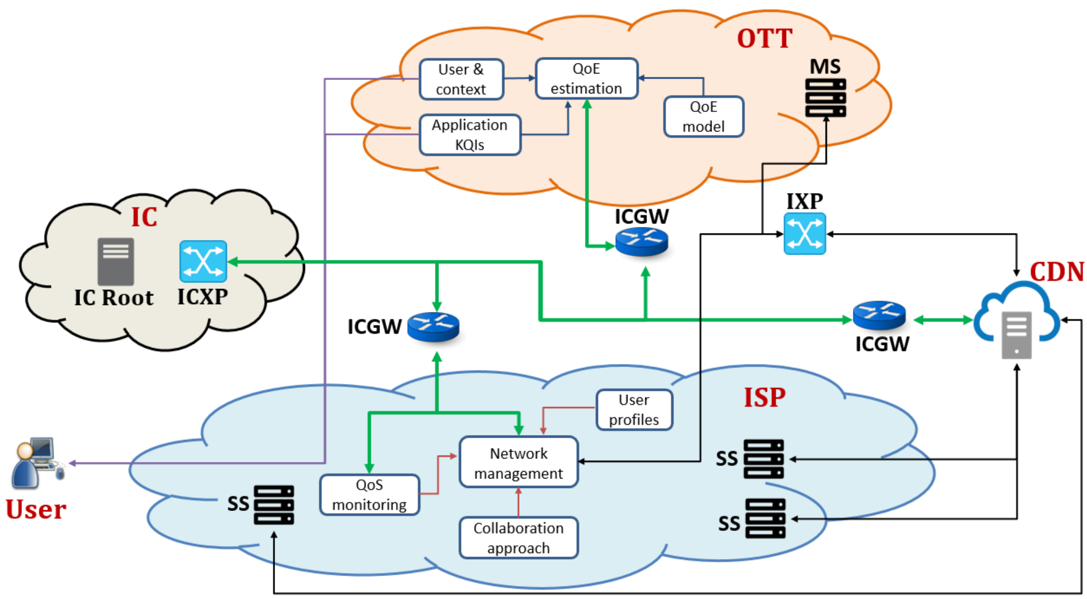

Fig. 1. Reference architecture for collaboration among ISP and OTTs.

state, that is not easily accessible by the OTT. The measured QoE is then used by the OTT to drive application level quality optimizations at the media server side and/or distribution of the contents with CDNs to lower the distance between the contents and the end users. However, in case of no collaboration, the SSs are typically placed at the edge, but outside, of the ISP network; instead, with the proposed collaboration, the ISP places the SSs inside its local network so that the content retrieval time is minimized.

The role of the ISP is mainly QoS monitoring and optimization of network resource allocation. Additionally, with the proposed collaboration the ISP provides settlement-free peering interconnection and also hosts the SSs of the collaborating OTTs in its network. With such a collaboration, the ISP reduces the transit costs; indeed, if contents are provided by direct peering and SSs, the ISP does not have to pay transit providers. Moreover, the provision of better service's quality to the end users may decrease the user churn for both the providers. The Network management module of the ISP is the core module that decides how the network resources should be managed as a function of the considered collaboration approach, which could be one of those discussed in Section 5. These collaboration approaches require some information to be shared between the ISP and the OTTs, so that the QoE can be accurately predicted and monitored to drive the network management. Different types and amounts of information may be shared, depending on the service provided as well as on the collaboration agreement: however, even the sharing of a minimal, but critical, amount of information may aid the ISP in making QoE-aware network management decisions, as opposed to the case of no collaboration.

Contents and data are exchanged at the IXPs, which are neutral physical infrastructures where the ASs (ISP, OTTs and CDN in this case) can exchange (mostly for free) Internet traffic with the others ASs with which they have a peering agreement. On the other hand, to allow the sharing of information and services among the ASs, we aim to base on the concept of the InterCloud (IC), motivated by the current *Softwarization* trend that has brought most of the biggest OTTs (e.g., Microsoft, Amazon, Google, Facebook) to implement their own cloud systems to provide their services all over the world. Even the future of ISPs is on the software with the shift towards network virtualization technologies, such as Software Defined Networking (SDN) and Network Function Virtualization (NFV).

The IC paradigm is supported by the definition of specific protocols and format that allow to create a network of clouds where content, storage, and computing functionalities are ubiquitous and interoperable [9]. The IC is composed of three main elements [33], which are shown in Figure 1: i) the *IC Root* is the core of the IC, which hosts the root servers containing all common mechanisms used by clouds participating to the IC, such as Naming Authority, Trust Authority, and Directory services. It acts as a broker in the IC overlay network and hosts a global cloud resource catalog which can be explored in order to discover desired cloud services; ii) the *IC eXchange* Points (ICXPs) are Neutral Access Points where clouds can interoperate and IC traffic can be properly handled (similarly to IXPs for network traffic). ICXPs provide negotiation and collaboration capabilities among heterogeneous and autonomous cloud environments. Each ICXP is affiliated with a particular IC Root element and hosts second-tier services; iii) the *IC Gateways* (ICGWs) are Internet routers representing interfaces between a particular cloud and the IC network. ICGWs translate IC requests and responses to the individual and customized protocol used by providers internally. Therefore, the IC addresses different problems related to cloud interoperability and specifically enables connectivity among disparate cloud environments avoiding the point-to-point collaboration (n^2 problem).

Being part of the IC, the ASs would be able to interoperate with each other for a combined QoE-aware management of network and services. For QoE monitoring, the OTT can make use of a cloud computing service to estimate in real-time the QoE of a specific service using QoE models (e.g., parametric models) specifically defined for the monitored service. A survey on parametric QoE estimation models for popular services is provided in [39]: such models are typically a function of both network KPIs and application KQIs. Being part of the IC, and establishing a collaboration agreement with the OTT, the ISP would be allowed to participate to the cloud computing service dedicated to QoE estimation to provide accurate information regarding the network KPIs, that the OTT alone could not be able to measure. The estimated QoE will be then the input of the Network management module run by the ISP, which in turn is a cloud computing service that takes decision on network management actions to target specific objectives depending on the collaboration approach between the providers.

To exchange information among the providers participating to the IC, we propose to use the XMPP protocol. Indeed, XMPP enables the near-real-time exchange of small pieces of structured data which are called *XML stanzas* [33]. XMPP specifies three types of XML stanzas: i) *message*, used to exchange messages among two entities over the XMPP overlay network; ii) *presence*, used to expose the availability of an entity in the XMPP overlay network; iii) *Info/Query (IQ)*, which allows for exchange request and response XML stanza. Besides stanza specific attributes, the XML root elements must also have addressing attributes, such as a *from* and a *to* attribute, whose values must contain a valid Jabber Identifier (JID) that in turn identifies the address of an XMPP entity. As an example, with reference to the aforementioned collaboration between OTT and ISP for QoE measurement, these would be the steps to be followed: i) the OTT uses the IQ stanza type `get` to explore the capabilities of the ISP's methods for QoS monitoring; ii) the ISP uses the IQ stanza type `result` to describe all QoS monitoring methods and actions that can perform; iii) the OTT can use common types defined for ISP's methods, i.e., `GET`, `POST`, `PUT`, and `DELETE`, which, together to the `request` and `response` elements (which describe the input and output representation data to be transferred), allow the OTT to make use of the ISP's needed methods for QoS monitoring. Also, the message stanza can be used to exchange information between the providers, such as the measured QoE, the type of service provided by the OTT, and/or particular user and context information, which may help the ISP to drive QoE-aware network management algorithms.

Table 2. Variables used in the article.

Variable	Description
$i = 1, 2, \ldots, I$	Index of the OTT services
$j = 1, 2, \ldots, J$	Index of the classes of service (CoS)
$n = 1, 2, \ldots, N$	Index of the common customers between the ISP and all the OTTs
$l = 1, 2, \ldots, L$	Index of the location zones (LZs) controlled by the ISP
R	Revenue for the provider
P	Price paid by the customer to subscribe to the service
$U(QoE)$	Customer churn utility function
ζ	Number of new customers who subscribe with marketing
α, β	Weights to determine provider's value for the CVL-based approach
F	Fairness index
ρ	Relevance factor to indicate the importance of fairness

An important point is what information is shared and how frequently. What information is shared depends on the collaboration agreement between the ISP and the OTT, so we refer to Section 5 where the collaboration approaches are presented. Since there are different types of information, we consider different updating frequencies for different types of information. We distinguish between static and dynamic information: *static information* does not change very frequently (weekly, monthly or even annually) and mainly regards some user information such as the user profile (end device, service subscription); this information can be updated only in case there are some changes. *Dynamic information* should be updated more frequently, especially when the user is active, because it regards service information such as application and context parameters as well as user's perceived QoE. Indeed, this information should be shared with the ISP to manage the network on the basis of service requirements, such as provide higher QoS in case the user's perceived QoE is too low. In order not to continuously exchange dynamic information, event-based techniques can be used so that information is exchanged only when specific events occur, such as quality degradation below specific thresholds [7], new services activated on the network, network congestion.

5 COLLABORATION STRATEGIES

In this section, we propose three collaboration approaches based on revenue/profit maximization and QoE fairness maximization. The first, (joint-venture, proposed in [6]), aims to maximize the revenue by providing better QoE to customers paying more. The second (CLV-based), aims to maximize the profit by providing better QoE to the most profitable customers (MPCs). The third (QoE-fairness-based), aims to maximize the QoE and the QoE fairness among all customers.

For each approach, we consider I OTT services (indexed by $i = 1, 2, \ldots, I$) that establish a collaboration agreement with an ISP to provide their services in J classes of service (CoS, indexed by $j = 1, 2, \ldots, J$). Also, we consider a total number N of customers that are in common among one (or more) of the OTTs and the ISP, i.e., the total number of users of the ISP's network. Table 2 lists the variables used in the proposed approaches.

5.1 Joint-venture approach

The joint-venture (JV) approach refers to the JV collaboration between OTTs and ISP proposed in our past work. With such a collaboration, OTTs and ISP offer their services jointly into different classes of service: the higher the price to subscribe to a CoS, the higher is the quality of the service provided to the subscribers of that CoS. In this section, we briefly describe the proposed approach but more details are found in [6].

The objective of the JV collaboration is to maximize the total revenue for both ISP and OTTs generated at time t_x (with $x = 0, 1, 2, \ldots, X$), namely R^x, which is defined as

$$R^x = \sum_{i=1}^{I} R_i^x = \sum_{i=1}^{I}\sum_{j=1}^{J} N_{i,j}^x P_{i,j} = \sum_{i=1}^{I}\sum_{j=1}^{J} \left(U_{i,j} N_{i,j}^{x-1} + \zeta_{i,j}\right) P_{i,j}, \tag{5}$$

where R_i^x is the combined revenue between the ISP and the i-th OTT, $N_{i,j}^x$ is the total number of customers subscribed to the j-th CoS of the i-th OTT at time t_x, $P_{i,j}$ is the price paid by the customers to subscribe to the j-th CoS of the i-th OTT, $U_{i,j}$ is the customer churn utility function for the j-th CoS of the i-th OTT, and $\zeta_{i,j}$ accounts for the customers subscribing to the j-th CoS of the i-th OTT with marketing and advertisement. The customer churn utility function computes the percentage of customers leaving/keeping the service as a function of the customer's perceived QoE

$$U_{i,j}(QoE_{i,j}) = \frac{1}{1 + e^{-z(QoE_{i,j} - QoE_{i,j}^m)}}, \tag{6}$$

where $QoE_{i,j}$ is the quality perceived by the customers subscribed to the j-th CoS of the i-th OTT, $QoE_{i,j}^m$ is the quality level at which half of the customers subscribed to the j-th CoS of the i-th OTT leave the service, and z is the sensitivity of the customers with respect to the paid price: customers who pay more expect higher quality [4].

Algorithm 1 describes the steps to be followed for the JV approach. The role of the OTT is to measure the average QoE perceived by the customers for each CoS ($\overline{QoE_{i,j}}$); then, this information is shared with the ISP, which has to maximize the revenue as a function of the QoE and the price paid by the customers so that customers who pay more receive higher QoE resulting in higher revenue for the providers. The average revenue is maximized during a reference period $t_{X_1} - t_{X_2}$ with the following equation

$$\bar{R}^* = \max_{\overline{QoE_{i,j}}, P_{i,j}} \left[\sum_{x=X_1}^{X_2} \left(\frac{\sum_{i=1}^{I} (\sum_{j=1}^{J} (U_{i,j} N_{i,j}^{x-1} + \zeta_{i,j})) P_{i,j}}{t_{X_2} - t_{X_1}} \right) \right]. \tag{7}$$

Therefore, the ISP manages its network in order to provide better QoS to the customers subscribed to higher-price CoS with the objective to maximize the QoE perceived by these customers and then reduce the customer churn for these CoS. Network traffic management techniques such as DiffServ, traffic prioritization, best routing path, are used by the ISP in this case. Although how the OTT and the ISP decide to divide the revenue is out of the scope of this article, a possibility may be to share the revenue proportionally to the amount of investment made by the providers over the total amount of investment for the service delivery. However, the criterion to be used in this scenario will be mostly based on political discussions among the different parties involved.

5.2 CLV-based approach

The JV approach presented in Section 5.1 is based on revenue maximization. However, as discussed in Section 2.2, the revenue just depends on the price paid by the customers to subscribe to the services, but it makes no consideration of the service profitability as well as of the cost of managing the customer service within the network. Therefore, it is important for the providers to understand the *value* of each customer in terms of profit, i.e., the revenue minus the costs. We identified the Customer Lifetime Value (CLV) as the metric for selection of the most profitable customers (MPCs) with the collaboration, i.e., those customers that for the providers have more value than others. Customers that are not MPCs are defined as standard customers (SCs). The CLV may be a function

ALGORITHM 1: Joint-venture approach

```
On the OTT side
for ∀OTT_i, i ∈ I do
    for ∀CoS_j, j ∈ J do
        OTT_i computes the average QoE_{i,j}, \overline{QoE_{i,j}}, perceived by the customers of the j-th CoS
        OTT_i computes the customer churn U_{i,j} as a function of \overline{QoE_{i,j}}
        OTT_i takes into account the number of new customers ζ_{i,j}
        OTT_i computes the current number of customers N^x_{i,j}
    end
    OTT_i computes the revenue R^x_i
    OTT_i shares with the ISP the \overline{QoE_{i,j}} and R^x_i values
end
On the ISP side
for ∀OTT_i, i ∈ I do
    for ∀CoS_j, j ∈ J do
        ISP computes the total revenue R^x with Eq. (5)
        ISP maximizes \bar{R}^* over \overline{QoE_{i,j}} and P_{i,j} with Eq. (7)
    end
end
```

of several parameters, such as the revenue provided by the customer, the number of loyal years and the profit margin provided to the company. Also customer's information may be taken into account to determine the CLV, such as laziness to churn (sleepers customers) or criticality (customers who often change subscription): in this case, low CLV may be assigned to sleepers whereas high CLV may be assigned to critical customers. Therefore, how the CLV is computed is strictly related to the needs of the providers and does not affect the validity of the proposed approach whose objective is to provide high quality to those customers that are the most valuable for the providers to avoid their churn. Customers with the highest values of CLV are considered as the MPCs and should be subjected to retention actions because the cost of winning a new customer is typically much larger than the cost of retaining an existing one [29], and the retention of the MPCs is even more convenient. As demonstrated in [6], the higher the QoE the lower the customer churn; therefore, the best way to decrease the churn of the MPCs is to provide them the best possible QoE.

We assume that ISP and OTTs have their own methods to compute the CLV (it may vary from company to company) but they just provide normalized CLV data not to disclose the real absolute values to the external world still allowing the collaboration to work properly. By considering the total number of common customers N, let CLV_n^{ISP}, $CLV_{n,i}$, and CLV_n^{OTT} be respectively the CLV of the n-th customer computed by the ISP, the i-th OTT and all the OTTs (if the customer is subscribed to more than one OTT). Then, we express the total CLV of the n-th customer as

$$CLV_n^{TOT} = \alpha CLV_n^{ISP} + (1-\alpha) CLV_n^{OTT} = \alpha CLV_n^{ISP} + (1-\alpha) \sum_{i=1}^{I} \beta_i CLV_{n,i}, \tag{8}$$

where α is a weight that determines the value of ISP and OTTs' customers for the total CLV computation, while β_i is a weight that determines the value of the OTTs in the collaboration agreement with the ISP: such weights are decided by ISP and OTTs on the basis of their agreement conditions. The last term takes into account the case of the same customer subscribed to more than one OTT: $\beta_i > 0$ when subscribed to the i-th OTT, otherwise $\beta_i = 0$. The N customers are then

ALGORITHM 2: CLV-based approach

```
On the OTT side
for ∀OTT_i, i ∈ I do
    for ∀ Customer n ∈ N do
        OTT_i computes CLV_{n,i} and QoE_{n,i}
    end
    OTT_i shares the information of CLV_{n,i} and QoE_{n,i} with the ISP
end
On the ISP side
for ∀ Customer n ∈ N do
    ISP computes CLV_n^{ISP}, CLV_n^{OTT} and CLV_n^{TOT}
    ISP sorts the N customers according to CLV_n^{TOT} to define the MPCs and SCs
    ISP sorts the L LZs according to the number of the MPCs
    for ∀LZ_l , l ∈ L do
        ISP places the SSs in the LZs where there is the highest number of the MPCs
    end
end
```

sorted as a function of CLV_n^{TOT}: the higher the CLV_n^{TOT}, the most profitable is the customer for both the ISP and the OTTs.

The objective of the ISP-OTT collaboration is to implement a QoE-aware management of the network to minimize the churn of the customers with the highest CLV_n^{TOT} so as to maximize the profit of both the providers. Firstly, the N customers are classified into two classes depending on their CLV_n^{TOT}. Specifically, a threshold is chosen to distinguish between the MPCs and the SCs, which depends on the difference of CLV_n^{TOT} among the customers. However, since resources are limited and costly, the percentage of MPCs should not be too high, such as the 20% or 30% for example. Though, these decisions should be evaluated case by case. Secondly, within the whole territory controlled by the ISP are identified the location zones (LZs) where most of the MPCs are situated. Let $l = 1, 2, \ldots, L$ indexes the LZs controlled by the ISP: these LZs are ranked on the basis of the number of MPCs belonging to each of them. Following this ranking, more network and application resources are provided to the highest ranked LZs to optimize the QoE of these customers and lose as few MPCs as possible.

Since the focus is on multimedia services, for QoE optimization several actions may be possible, such as i) network management implemented by the ISP for traffic prioritization on the basis of the type of service to be delivered [31]; ii) the places where the SSs should be located must be as close as possible to the highest ranked LZs; iii) the contents replicated on the SSs must be the most viewed by the MPCs. Indeed, as discussed in [27], CDN basically deals with the issues of identifying where to place the SSs and identifying what content to replicate on the servers. With the proposed collaboration approach, these issues are solved with a major focus on profit maximization. In fact, if the SSs will be situated closer to the highest ranked LZs, the MPCs belonging to these LZs will perceive better QoE due to the improvement of service delivery.

Algorithm 2 describes the steps to be followed for the CLV-based approach.

5.3 QoE-fairness-based approach

The collaboration approaches proposed in Sections 5.1 and 5.2 are based respectively on revenue/profit maximization by providing better QoE to customers paying more or which are the most profitable for the providers. However, one of the ultimate goals in future multimedia networks is to provide a fair user-centric quality, so that the user QoE is maximized for all users in a network. The reason is that selfish applications may maximize its own QoE, potentially worsening QoE levels of users of different applications. According to the general fairness metric defined in [21], which satisfies QoE-relevant properties, a system is defined absolutely QoE fair when all users receive the same QoE value. Therefore, the objective of the proposed collaboration scenario is to maximize the QoE and the QoE fairness of all the N customers in common among the ISP and the OTTs. Specifically, we consider maximization of QoE and QoE fairness within each CoS $j \in J$ because customers who pay more expect higher quality [4].

With regard to the QoE fairness index, F, we refer to that defined in [21], as follows

$$F = 1 - \frac{\sigma}{\sigma_{max}}, \tag{9}$$

where σ and σ_{max} are respectively the standard deviation and the maximum possible standard deviation of the QoE levels which may be maximally perceived by the users. Then, the higher the variance among the QoE values perceived by different users, the lower the fairness. Also, the fairness value depends on the considered quality scale because $\sigma_{max} = \frac{H-L}{2}$, where H and L are respectively the maximum and minimum values of the QoE evaluation scale. For example, considering the ITU 5-level quality scale so that $L = 1$ and $H = 5$, Eq. (9) becomes

$$F = 1 - \frac{2\sigma}{5-1} = 1 - \frac{\sigma}{2}. \tag{10}$$

Therefore, considering the 5-level quality scale, the QoE fairness index computed among all users of the j-th CoS of the i-th OTT, $F_{i,j}$, is

$$F_{i,j} = 1 - \frac{\sigma_{i,j}}{2} = 1 - \frac{1}{2}\sqrt{\frac{\sum_{n=1}^{N_{i,j}} (QoE_{n,i,j} - \overline{QoE_{i,j}})^2}{N_{i,j} - 1}}, \tag{11}$$

where $\sigma_{i,j}$ is the standard deviation of the QoE levels perceived by the customers subscribed to the j-th CoS of the i-th OTT, $QoE_{n,i,j}$ is the QoE perceived by the n-th customer subscribed to the j-th CoS of the i-th OTT, and $\overline{QoE_{i,j}}$ is the average QoE perceived by all $N_{i,j}$ customers subscribed the j-th CoS of the i-th OTT.

On the other hand, the QoE fairness of the network, computed among all the N_j customers of the ISP's network subscribed to the j-th CoS of all the OTTs, F_j^{ISP}, is defined as

$$F_j^{ISP} = 1 - \frac{\sigma_j^{ISP}}{2} = 1 - \frac{1}{2}\sqrt{\frac{\sum_{i=1}^{I} \sum_{n=1}^{N_{i,j}} (QoE_{n,i,j} - \overline{QoE_j})^2}{N_j - 1}}, \tag{12}$$

where σ_j^{ISP} is the standard deviation of the QoE levels perceived by all customers subscribed to the j-th CoS of all the OTTs, $QoE_{n,i,j}$ is the QoE perceived by the n-th customer subscribed to the j-th CoS of the i-th OTT, and $\overline{QoE_j}$ is the average QoE perceived by all N_j customers subscribed to the j-th CoS of all the OTTs.

The objective is to maximize the fairness index, F_j^{ISP}, as well as the average QoE of all customers, $\overline{QoE_j}$, within each CoS. To this, we refer to the maximum utility function proposed in [17]

$$U_j = (1-\rho)\overline{QoE_j} + \rho F_j^{ISP}, \tag{13}$$

where ρ is the relevance factor that indicates the importance of fairness in the management decision. On the one hand, the OTTs predict the QoE perceived by their customers using QoE metrics that may take into account application, human and context influence factors, and try to maximize this QoE by making the best possible use of the available network resources. On the other hand, OTTs do not have any control over the network infrastructure, whose performance has a strong influence on QoE and is shared among multiple and different OTT services. Therefore, the collaboration with

ALGORITHM 3: QoE-fairness-based approach

On the OTT side
for ∀ $OTT_i, i \in I$ **do**
  for ∀ $CoS_j, j \in J$ **do**
    for ∀ $Customer\, n \in N$ **do**
      OTT_i computes the QoE perceived by the $N_{i,j}$ customers subscribed to the j-th CoS, $QoE_{n,i,j}$
    end
    OTT_i shares with the ISP the information of $QoE_{n,i,j}$
  end
end
On the ISP side
– ISP computes the average QoE perceived by all the N_j customers of the j-th CoS, $\overline{QoE_j}$
– ISP manages the network for the combined maximization of $\overline{QoE_j}$ and F_j^{ISP} using Eq. 13

the OTTs, that share the computed QoE of the customers, allows the ISP to globally manage the network for the combined maximization of the overall QoE and QoE fairness. The value of the relevance factor ρ is decided by ISP and OTTs depending on their agreements.

Algorithm 3 describes the steps to be followed for the QoE-fairness-based approach.

5.4 Limits and assumptions

Each of the proposed three models is based on some assumptions and has some limits. The JV approach assumes that OTT and ISP jointly provide the service to the customers, which have to paid with a single subscription. The main limit is to define how OTT and ISP should divide this revenue because it is not easy to translate the information and effort provided by the providers in terms of money. From the technical point of view, the network should be able to give priority to the traffic of the priority classes, which however is a reasonable assumption considering the currently available QoS technologies. The CLV-based approach assumes that the providers measure the value of each customer as a function of some specific parameters. The main limit is that a customer that is profitable for the ISP (OTT) may not be profitable for the OTT (ISP or some other OTTs). Therefore, how all customers should be ranked in terms of total profitability should be accurately defined by the providers not to fail to meet some provider's collaboration expectations (in terms of profit). Finally, the QoE-fairness-based approach is based on the provision of fair QoE levels to customers of different services. This approach has the limit that the utilization of network resources may violation the network neutrality because network QoS is not linearly related to the user's perceived QoE. As an example, the amount of network resources needed to provide a high quality video streaming service are typically more than those needed to provide a high quality VoIP or web browsing service. It is however important to note that also the other two techniques are violating somehow the network neutrality principle.

6 SIMULATIONS-BASED ANALYSIS

The simulations are performed to analyze the effectiveness of the aforementioned collaboration strategies in terms of the QoE delivered to the users, QoE fairness among the users and providers' profit. Section 6.1 provides the details of the simulation setup whereas in Section 6.2 simulation results are shown and discussed.

Table 3. Simulation parameters and settings.

Parameters	Settings
Initial total No. of VoIP and video streaming users	100
Initial No. of PR and ST users in VoIP and video streaming	50
Price paid by PR users	0.6
Price paid by ST users	0.3
Cost associated with PR users	$0.1 - 0.5$[1]
Cost associate with ST users	$0.1 - 0.2$[1]
No. of loyal years	$1 - 5$[1]
Sensitivity of the user with the price, z	3.5
Quality level at which half of ST users leave the service, $QoE^{m}_{i,ST}$	2.5
Quality level at which half of the PR users leave the service, $QoE^{m}_{i,PR}$	3.5
Weights for the CLV-based approach, α, β	0.5

6.1 Simulation setup

The simulations are conducted on the MATLAB platform. We considered two different OTT services: VoIP and video streaming. The aforementioned applications are selected according to the studies conducted in [15, 26], where video streaming and VoIP are considered as the most sensitive multimedia applications with reference to network resources utilization. The setting of the major parameters have been done as follows. Table 3 summarizes simulation parameters and settings.

CoS, prices and costs: For each of the proposed approaches, two different CoS are considered, on the basis of the users' willingness to pay: 1) Premium (PR) service and; 2) Standard (ST) service. The price to subscribe to the PR service is higher than that for the ST service, because the PR service should guarantee a higher quality experience to the users. As a consequence, the providers may incur higher costs for the provision of the PR service than for the ST service, which is provided through the best effort approach. Prices and costs are normalized in the $0 - 1$ scale. We set the price to be paid by PR and ST users respectively to 0.6 and 0.3, on the basis of our previous studies, where maximum revenue was generated on these prices [6]. Additionally, we consider the cost associated to PR and ST users to range randomly between 0.1 and 0.5, and between 0.1 and 0.2, respectively, so that a minimum profit of 0.1 is always assured to the providers. We consider the cost to vary from user to user because many causes may infer service costs, such as the distance of the user from the server. Note that for the CLV-based approach, both PR or ST users may be considered as MPCs or SCs, because they are further classified on the basis of their CLV value.

Initial population and churn: We consider a starting number of 100 users for both VoIP and video streaming applications, with 50% of the users belonging to the PR service and 50% to the ST service. Each user is assigned with the following characteristics: 1) price paid by the user and cost associated to keep the user in the service, which depend on the CoS the user belongs to; 2) profit associated with the user, calculated as the price minus the cost and; 3) number of loyal years of the user in the service, which are assigned randomly in the range $1 - 5$ years. In order to make the fair comparison among the collaboration strategies, the same initial pole of users are considered for each strategy. The user churn is calculated using the Eq. (6), where $z = 3.5$ for each CoS while $QoE^{m}_{i,j}$ is selected to be 2.5 and 3.5 for ST and PR users, respectively. Note that $QoE_{i,j}$ is computed as the average QoE of all users' QoE predictions per month. On the basis of the computed user churn, the number of the users in each CoS and type of application are updated on the monthly

[1]These parameters are randomly assigned with uniform distribution.

basis. We also consider the number of users joining the service each month $\zeta_{i,j}^{x}$, which is computed as a random number from the Poisson distribution with the mean equal to the 5% of users belonging to each CoS of an application. A random cost is assigned to each new user, depending on the CoS the user belongs to. Furthermore, the fairness index in each CoS is computed on the monthly basis using Eq. (12).

Quality models: We consider the E-model for the VoIP application, which is a planning parametric model defined by the ITU for VoIP applications, which measures the quality in terms of the R-factor ranging from 0 to 100, where 100 is the best quality [22]. In our simulations, we use the simplified version of the E-model, proposed in [11], which emphasizes the effect of sources of quality degradation observed over data networks, namely one-way delay, packet loss ratio, and coding scheme. Furthermore, we considered [11] for converting the R-factor with values between 1 and 5 as the MOS. For the video streaming service, the QoE is predicted by using the model proposed in [18, 19] because it is particularly designed for the non adaptive HTTP video streaming

$$MOS = \alpha \cdot e^{-\beta(L) \cdot N} + \gamma, \tag{14}$$

where L and N are the stalling duration and the total number of stalling events during the video session, respectively. α, γ and $\beta(L)$ are some formula parameters that have been set as follows: $\alpha = 3.5$, $\gamma = 1.5$ and $\beta(L) = 0.15 \cdot L + 0.19$. Moreover, for a 30-second-long video session in [18], L and N are defined as a function of the buffer size d of the client player, the frame arrival rate λ and the video frame rate μ, such that $L = \frac{d}{\mu \cdot a}$ and $N = \frac{1-a}{d^*}$, whereas $a = \frac{\lambda}{\mu}$ and $d^* = \frac{d}{\mu}$ are the load factor and the normalized buffer size respectively. We consider the buffer size equal to 4 s as recommended in [18] for the video streaming case. In the simulations, network throughput and video bit rate are considered respectively as λ and μ (instead of frame arrival rate and video frame rate), as also considered in the model in [18].

Quality as a function of the number of customers: To understand the relationship between the QoE and the number of users we have used the network emulator Mininet. It is important to highlight that this analysis has been conducted offline with respect to the major MATLAB simulation framework. Indeed, we specifically needed the curves of quality versus the number of customers for the considered network. In the emulator, we considered the ISP's network to be composed of different virtual slices with a fixed channel capacity for each different CoS. For the VoIP service encoded with G.729a, the number of UDP flows with the encoding bit rate (equal to 31.2 kbps) are varied from 50 to 100 over a fixed capacity link of 1.95 Mbps. The mean playout delay and jitter are observed by varying the number of flows for both PR and ST VoIP use cases. The packet loss probability is computed from the cumulative density function of the observed mean playout delay and jitter as highlighted in [35]. The measured playout delay and packet loss probability are used for the computation of the R-factor [22] and delivered QoE in terms of MOS [11]. With regard to video streaming, the videos were encoded with the H.264 codec. The PR users stream 1080p HD in 1920×1080 resolution while the ST users stream 480p SD videos in 854×480 resolution; the frame rate for both HD and SD videos is 30 fps. The encoding bit rate for HD and SD videos varies from 3 Mbps to 6 Mbps and 500 Kbps to 2 Mbps, respectively.[4] For the video streaming, the traffic composed of TCP flows is generated where the number of flows is varied from 50 to 100 over fixed capacity links of 375 Mbps and 125 Mbps for PR and ST CoS respectively. The observed change in the throughput with the increase in number flows/users is used to measure the delivered QoE from Eq. (14). With this setting we wanted to stress the network with also sending traffic quite exceeding the network capacity. We would like to highlight that this setting an extensive modelling

[4]https://support.google.com/youtube/answer/2853702?hl=en

of quality versus the number of customers is not the main focus of the paper but it was important to conduct the analysis of the proposed algorithms performance.

Performed simulations: With the described setting, we ran the simulations that implemented the three proposed algorithms in MATLAB and we observed the evolution of the system, i.e., number of customers per class and service and the quality provided, over time providing results as the average over 100 runs. Each approach manages the network resources differently: the JV approach provides more network resources to PR users, who are the users paying more, to maximize the revenue (see Section 5.1); the CLV approach provides more network resources to the Most Profitable Customers (MPC) to maximize the profit (see Section 5.2); the QoE-fairness-based approach provides fair network resources to maximize QoE and QoE fairness among users of the same CoS (see Section 5.3). Again, to provide a fair comparison of the collaboration approaches, we assume that the considered ISP's network portion, where we apply the CLV-based algorithm, is the LZ where there is the highest number of MPCs. So we are not considering multiple LZs in the simulations.

6.2 Simulations results

This section provides the simulation results regarding the comparison of the three collaboration strategies, presented in Section 5, in terms of quality provided to the users, QoE fairness among the users and profit generated for the providers.

Figure 2 shows the average QoE delivered in each CoS of both video streaming and VoIP applications over a period of 2 years for each collaboration strategy. In case of the PR VoIP and video streaming, the CLV-based approach provides better quality as compared to JV and QoE-fairness-based approaches, i.e., 4.38 and 4.74 for PR VoIP and PR video streaming, respectively. The reason is that the number of users classified as MPCs (and are treated with PR CoS) on the basis of CLV (30%) is less that those of the JV and QoE-fairness-based approaches (50% are PR users). Hence, better quality is delivered to MPCs in case of CLV-based approach because more resources are available for less users. However, in case of ST VoIP and video streaming, the CLV-based approach provides lower quality, i.e., 3.23 and 3.22 as compared to JV and QoE-fairness-based approaches because more users are classified as SCs (which are treated as ST users) on the basis of their CLV. Hence, lower quality is perceived by the SCs for VoIP and video services. Moreover, the JV approach appears to deliver better quality for PR video as compared to QoE-fairness-based approach.

Figure 3 provides the comparison in terms of the QoE fairness among the different collaboration strategies. The QoE fairness is computed within the two CoS, PR and ST, considering the QoE measured for the customers of both the video streaming and VoIP services. As expected, the QoE-fairness-based approach provides higher degree of fairness in terms of QoE among the same CoS. In case of PR VoIP and PR video streaming, the CLV-based approach provides more fairness at application level as compared to JV because the number of MPCs is lower than that of PR users of JV, hence, higher (and fair) quality is provided to the MPCs. However, in case of ST VoIP and ST video streaming, JV provides better fairness in comparison with the CLV-based approach for the opposite reason. Indeed, there are less ST users in JV than SCs in the CLV-based approach.

Figures 4-6 show the evolution in the number of users for the different approaches by the end of every month over a period of 2 years. The 95% confidence interval is shown, which is computed for the 100 runs executed. With regard to the JV approach, the number of ST and PR users increases with the time. However, the number of PR users has an important increase whereas the number of ST users has a slight increase reaching a saturation level. This is due to the lower quality perceived by the ST users (which pay less than the PR users but do not have a good guaranteed quality) that increases the number of churners. The number of users for the CLV-based evolves differently than that for the JV approach. Indeed, while for the JV approach 50% of the users paying less are

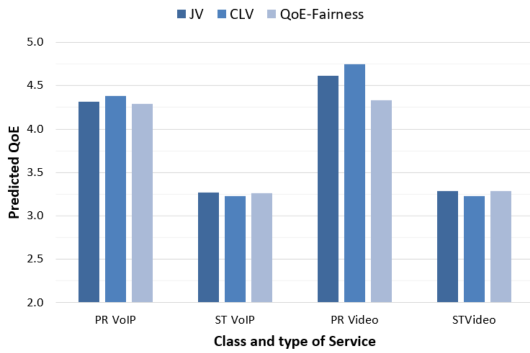

Fig. 2. Comparison in terms of delivered QoE.

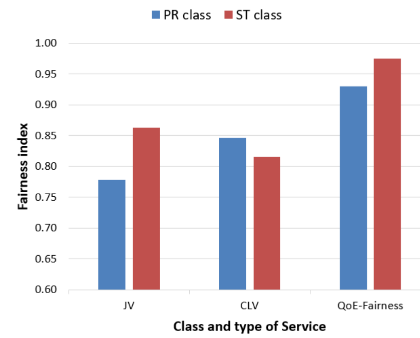

Fig. 3. Comparison in terms of QoE fairness.

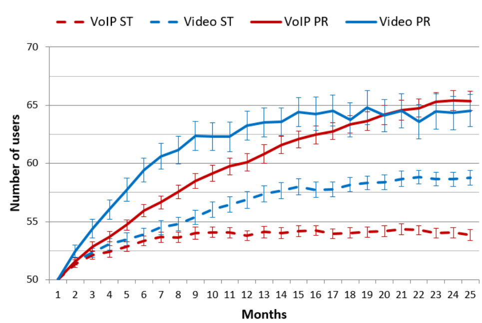

Fig. 4. Evolution in the number of users for the Joint Venture (JV) approach.

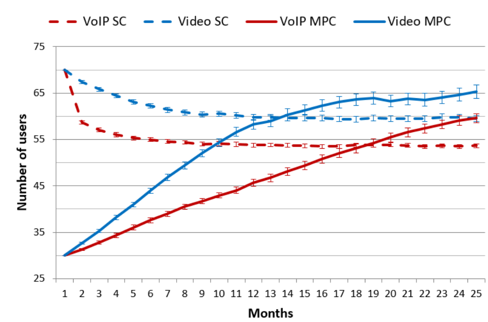

Fig. 5. Evolution in the number of users for the CLV-based approach.

considered ST users (50 users), for the CLV-based approach the users are further classified on the basis of CLV and mostly (70%, i.e., 70 users) are considered as SCs. Then, at the initial time most of network resources are already allocated for the SCs, who receive low quality services and leave the service. However, after some months a saturation level is reached. The opposite situation concerns the MPCs, which are lower than the SCs at the initial time, but have an important grow over the time due to the greater amount of network resources available which allow them to receive good service quality. Finally, with regard to the QoE-fairness-based approach, after an initial time in which the number of users for each CoS and service increases, the evolution of the number of users becomes more unpredictable and indented. This may be due to the fairness algorithm that, trying to provide fair QoE to each user, may provide insufficient network resources that in turn transform into insufficient QoE to the users, which become churners. Then, more network resources are available that allow new users to join. Again, the action of the algorithm may decrease the QoE causing new churners and so on. We remind that the value of ρ set for these simulations in Eq. (13) was $\rho = 0.9$ so as to apply a fairness-aware solution.

Finally, Figure 7 compares the collaborative approaches on the basis of the total profit generated by the providers over a period of two years. The profit per user is computed as the difference between the price paid by the user to subscribe to the CoS and the cost associated to keep the user into that CoS. The total profit is calculated as the sum of the profit of each user of both PR and ST CoS and both VoIP and video streaming applications. As expected, the CLV-based approach provides the highest profit, followed by the JV approach and the QoE-fairness-based approach.

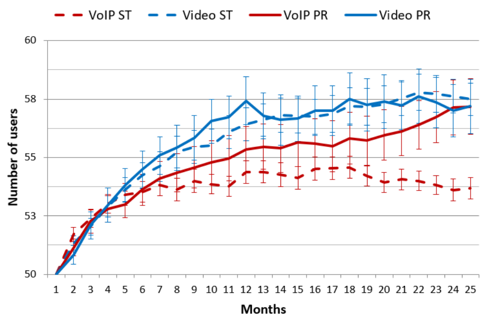

Fig. 6. Evolution in the number of users for the QoE-fairness-based approach.

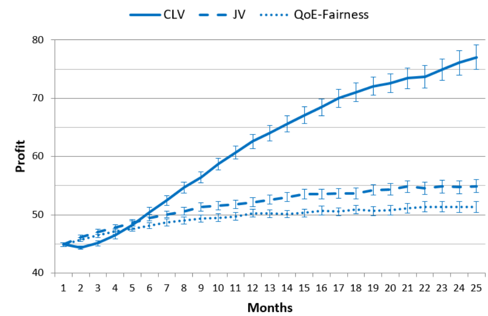

Fig. 7. Comparison in terms of the profit.

Differently from the JV approach, that provides better quality to users just paying more, the CLV-based approach considers the profit provided by each customer, by taking into account the cost associated to each user in addition to the price paid. Therefore, the CLV-based approach is able to provide more resources and then higher quality to the users who are the most profitable for the providers and then the total profit is higher with respect to the other two approaches. For example, it may happen that some users classified as PR by the JV approach are actually not profitable because their associated cost is high and they are customers just since one year (for example, $Price = 0.6$, $Cost = 0.5$, $Profit = 0.1$, $years = 1$, $CLV = 0.6 \times 0.1 \times 1 = 0.06$). On the contrary, some users classified as ST may be more profitable because their associated cost is low and they are customers from 5 years (for example, $Price = 0.3$, $Cost = 0.1$, $Profit = 0.2$, $years = 5$, $CLV = 0.3 \times 0.2 \times 5 = 0.3$). In this way, higher quality is provided not only to the most profitable customers but also to the loyal customers, and maybe even better results may be obtained by this approach if a model which takes into account also user's loyalty would be used (we based on the user churn model which defines if the user is satisfied just on the basis of the perceived quality and the price paid, loyalty is not modelled). Finally, as also discussed before, the QoE-fairness-based approach is mostly based on providing fair QoE to the users. For this reason, sometimes insufficient QoE level is provided to the users which become churners decreasing the profit of the providers.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we considered the collaboration among ISPs and OTTs for a joint QoE-aware service management in terms of technical and economic aspects. Accordingly, we proposed a reference architecture for a possible collaboration and information exchange among them. We also proposed three different collaboration strategies, namely: Joint-venture, CLV-based and QoE-fairness-based approaches. Then, we conducted simulations to compare the performance of these approaches in terms of quality delivered to the users, QoE fairness among users and total profit generated by providers. From simulation results, the CLV-based and JV approaches provide better QoE to PR and ST users, respectively. As expected, the QoE-fairness-based approach provides higher degree of QoE fairness among the users of the same CoS as well as on the network level. With regard to the profit, the CLV-based approach provides the best results, followed by the JV approach and the QoE-fairness-based approach.

In future works, we aim to improve the simulations of our proposed collaboration approaches by considering virtualized networks, such as Software-Defining Networks (SDN), to test the performance of the approaches on a virtual network environment. Indeed, with SDN it would be possible to run the algorithm on the SDN controller whereas on the network plane management actions can allocate network resources on the basis of the algorithm's decisions. Also, further studies are required to evaluate the impact of frequency of exchange of information among the OTTs and ISP on the network load as well as on the amount of data and the cost of data storage. Moreover, QoE models valid for longer time periods need to be studied in combination with a more accurate user churn function, which are required to drive the collaborative management of the services. Furthermore, more accurate models are required to correlate the change in the QoE and QoS with the change in the number of users/flows.

ACKNOWLEDGMENTS

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

REFERENCES

[1] 2000. A Framework for Integrated Services Operation over Diffserv Networks. (2000). NWG RFC 2998.
[2] 2012. Co-optimisation of overlay applications and underlying networks. (2012). ENVISION FP7 Project.
[3] 2014. Application-Layer Traffic Optimization (ALTO) Protocol. (2014). IETF RFC 7285.
[4] Imad Abdeljaouad and Ahmed Karmouch. 2015. Monitoring IPTV Quality of Experience in overlay networks using utility functions. *Journal of Network and Computer Applications* 54 (2015), 1–10.
[5] Accenture. 2017. Customer Service, Not Price, Remains Top Cause of Customer Churn, Accenture Study Finds. (2017). Retrieved November 23, 2017 from https://newsroom.accenture.com/subjects/customer-relationship-management/customer-service-not-price-remains-top-cause-customer-churn-accenture-study-finds.htm
[6] Arslan Ahmad, Alessandro Floris, and Luigi Atzori. 2016. QoE-centric service delivery: A collaborative approach among OTTs and ISPs. *Computer Networks* 110 (2016), 168–179.
[7] Arslan Ahmad, Alessandro Floris, and Luigi Atzori. 2017. Towards QoE monitoring at user terminal: A monitoring approach based on quality degradation. In *Proc. of the IEEE Int. Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2017.* IEEE.
[8] Sabina Baraković and Lea Skorin-Kapov. 2013. Survey and Challenges of QoE Management Issues in Wireless Networks. *Journal of Computer Networks and Communications* 2013 (2013).
[9] David Bernstein, Erik Ludvigson, Krishna Sankar, Steve Diamond, and Monique Morrow. 2009. Blueprint for the Intercloud - Protocols and Formats for Cloud Computing Interoperability. In *Proc. of the Fourth Int. Conf. on Internet and Web Applications and Services, 2009.* IEEE, 328–336. https://doi.org/10.1109/ICIW.2009.55
[10] Cisco. 2017. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper. (2017). Retrieved November 23, 2017 from https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html
[11] Robert G. Cole and Joshua H. Rosenbluth. 2001. Voice over IP performance monitoring. *ACM SIGCOMM Computer Communication Review* 31, 2 (2001), 9–24.
[12] FierceWireless. 2017. iPhone users cite video buffering as biggest annoyance. (2017). Retrieved November 23, 2017 from http://www.fiercewireless.com/europe/iphone-users-cite-video-buffering-as-biggest-annoyance
[13] DASH Industry Forum. 2016. DASH-IF Position Paper: Server and Network Assisted DASH (SAND). (2016).
[14] Nicolas Glady, Bart Baesens, and Christophe Croux. 2009. Modeling churn using customer lifetime value. *European Journal of Operational Research* 197, 1 (2009), 402–411.
[15] Gerardo Gómez, Javier Lorca, Raquel García, and Quiliano Pérez. 2013. Towards a QoE-driven resource control in LTE and LTE-A networks. *Journal of Computer Networks and Communications* 2013 (2013).
[16] Janny C. Hoekstra and Eelko K. R. E. Huizingh. 1999. The Lifetime Value Concept in Customer-Based Marketing. *Journal of Market-Focused Management* 3, 3 (1999), 257–274.
[17] Tobias Hoßfeld, P. E. Heegaard, Lea Skorin-Kapov, and M Varela. 2017. No silver bullet: QoE metrics, QoE fairness, and user diversity in the context of QoE management. In *Proc. of the Ninth Int. Conf. on Quality of Multimedia Experience (QoMEX), 2017.* IEEE.
[18] Tobias Hoßfeld, Christian Moldovan, and Christian Schwartz. 2015. To each according to his needs: Dimensioning video buffer for specific user profiles and behavior. In *IFIP/IEEE Int. Symposium on Integrated Network Management (IM), 2015.* IEEE, 1249–1254.
[19] Tobias Hoßfeld, Raimund Schatz, Ernst Biersack, and Louis Plissonneau. 2013. Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience. In *Data Traffic Monitoring and Analysis*, Ernst Biersack, Christian

Callegari, and Maja Matijasevic (Eds.). Springer Berlin Heidelberg, 264–301.

[20] Tobias Hoßfeld, Lea Skorin-Kapov, Yoram Haddad, Peter Pocta, Vasilios A. Siris, Andrej Zgank, and Hugh Melvin. 2015. Can context monitoring improve QoE? A case study of video flash crowds in the Internet of Services. In *Proc. of the IFIP/IEEE Int. Symposium on Integrated Network Management (IM), 2015*. IEEE, 1274–1277.

[21] Tobias Hoßfeld, Lea Skorin-Kapov, Poul E. Heegaard, and Martin Varela. 2017. Definition of QoE Fairness in Shared Systems. *IEEE Communications Letters* 21, 1 (2017), 184–187.

[22] ITU. 2015. The E-model: a computational model for use in transmission planning. Recommendation ITU-T G.107.

[23] Patrick Le Callet, Sebastian Möller, and Andrew Perkis. 2012. Qualinet White Paper on Definitions of Quality of Experience (2012). (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.

[24] Eirini Liotou, Dimitris Tsolkas, Nikos Passas, and Lazaros Merakos. 2015. Quality of Experience Management in Mobile Cellular Networks: Key Issues and Design Challenges. *IEEE Communications Magazine* 53, 7 (2015), 145–153.

[25] Nokia. 2017. Nokia White Paper, Quality of Experience (QoE) of mobile services: Can it be measured and improved? (2017). Retrieved November 23, 2017 from http://afutt.org/Qostic/qostic1/MOB-GD-MGQ-NOKIA-040129-Nokia-whitepaper_qoe_net-final.pdf

[26] Pablo Oliver-Balsalobre, Matías Toril, Salvador Luna-Ramírez, and José María Ruiz Avilés. 2016. Self-tuning of scheduling parameters for balancing the Quality of Experience among services in LTE. *EURASIP Journal on Wireless Communications and Networking* 2016, 1 (2016), 1–12.

[27] Andrea Passarella. 2012. A survey on content-centric technologies for the current Internet: CDN and P2P solutions. *Computer Communications* 35, 1 (2012), 1–32.

[28] Ramneek, P. Hosein, W. Choi, and W. Seok. 2015. Disruptive network applications and their impact on network neutrality. In *Advanced Communication Technology (ICACT), 2015 17th International Conference on*. IEEE, 663–668.

[29] Peter Reichl, Sebastian Egger, Sebastian Möller, Kalevi Kilkki, Markus Fiedler, Tobias Hoßfeld, Christos Tsiaras, and Alemnew Asrese. 2015. Towards a comprehensive framework for QoE and user behavior modelling. In *Proc. of the Seventh Int. Workshop on Quality of Multimedia Experience (QoMEX), 2015*. IEEE.

[30] Susanna Schwarzmann, Thomas Zinner, and Ognjen Dobrijevic. 2016. Towards a Framework for Comparing Application-Network Interaction Mechanisms. In *28th Int.l Teletraffic Congress (ITC 28), 2016*. IEEE, 13–18.

[31] Janne Seppänen, Martín Varela, and Aggeliki Sgora. 2014. An autonomous QoE-driven network management framework. *Journal of Visual Communication and Image Representation* 25, 3 (2014), 565–577.

[32] Michael Seufert, Sebastian Egger, Martin Slanina, Thomas Zinner, Tobias Hoßfeld, and Phuoc Tran-Gia. 2015. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys & Tutorials* 17, 1 (2015), 469–492.

[33] Alexander Stanik and Odej Kao. 2016. A proposal for REST with XMPP as base protocol for intercloud communication. In *Proc. of the 7th Int. Conf. on Information, Intelligence, Systems & Applications (IISA), 2016*. IEEE.

[34] StreamingMediaBlog.com. 2017. Chart Shows Which Content Owners Have Direct Interconnect Deals With ISPs. (2017). Retrieved November 23, 2017 from http://blog.streamingmedia.com/2014/05/chart-shows-which-content-owners-have-direct-interconnect-deals-with-isps.html

[35] Lingfen Sun and Emmanuel Ifeachor. 2004. New models for perceived voice quality prediction and their applications in playout buffer optimization for VoIP networks. In *IEEE Int. Conf. on Communications, 2004*, Vol. 3. IEEE, 1478–1483.

[36] Maria Torres Vega, Decebal Constantin Mocanu, Jeroen Famaey, Stavros Stavrou, and Antonio Liotta. 2017. Deep Learning for Quality Assessment in Live Video Streaming. *IEEE Signal Processing Letters* 24, 6 (2017), 736–740.

[37] Maria Torres Vega, Decebal Constantin Mocanu, and Antonio Liotta. 2017. Unsupervised deep learning for real-time assessment of video streaming services. *Multimedia Tools and Applications* 76, 21 (2017), 22303–22327.

[38] Maria Torres Vega, Decebal Constantin Mocanu, Stavros Stavrou, and Antonio Liotta. 2017. Predictive no-reference assessment of video quality. *Signal Processing: Image Communication* 52 (2017), 20–32.

[39] Dimitris Tsolkas, Eirini Liotou, Nikos Passas, and Lazaros Merakos. 2017. A survey on parametric QoE estimation for popular services. *Journal of Network and Computer Applications* 77, 1 (2017), 1–17.

[40] Martín Varela, Patrick Zwickl, Peter Reichl, Min Xie, and Henning Schulzrinne. 2015. From Service Level Agreements (SLA) to Experience Level Agreements (ELA): The challenges of selling QoE to the user. In *Proc. of the IEEE Int. Conf. on Communication Workshop (ICCW) 2015*. IEEE, 1741–1746.

Received June 2017; revised November 2017