MNO-OTT Collaborative Video Streaming in 5G: The Zero-rated QoE Approach for Quality and Resource Management

Arslan Ahmad and Luigi Atzori
IS-Wireless, Poland. Email: a.ahmad@is-wireless.com
DIEE, University of Cagliari, Italy. Email: l.atzori@ieee.org

***Abstract*—The Quality of Experience (QoE) management procedures for multimedia services benefit from an effective collaboration between the Mobile Network Operators (MNOs) and the Over-The-Top (OTT) service providers as the former can allocate the appropriate network resources to the users and the latter has access to key influence factors for having a proper view of the provided QoE. One successful collaboration model is the *zero-rated data rate approach*, according to which the MNO limits the data rate of the users towards the collaborating OTT applications with the benefit for the user that the generated traffic is not counted in her monthly contract data limit. Accordingly, the MNO may reduce the network congestion, and the users are encouraged to select the collaborating OTT applications. Though, this approach does not consider the resulting QoE, which may vary significantly from one user to another even if the same throughput is given.**

Based on this consideration, in this paper, we proposed the *zero-rated QoE approach*, according to which the limit is introduced in terms of QoE rather than throughput. This clearly requires a stronger collaboration between the OTT and the MNO, as the first has to give access to the second to quality influence factors and the latter to allocate resources according to the predicted QoE. The contributions of this paper are: the introduction of the novel zero-rated QoE approach with particular reference to video streaming services; the definition of novel components in the 3GPP architecture so that this approach can be introduced; the definition of the algorithm for the allocation of the appropriate radio resources to each user; a simulation analysis where the proposed approach is compared with respect to the former zero-rated approach, which shows significant improvements in terms of average provided quality and quality fairness at the same overall throughput.

Index Terms—5G, Quality of Experience (QoE), QoE management, Video streaming, Over-The-Top (OTT), Mobile Network Operators (MNOs), Radio resource management.

I. Introduction

TODAY, Internet traffic mostly consists of multimedia traffic due to the popularity of Over-The-Top (OTTs) services such as YouTube, Netflix, and Facebook. Among all the multimedia services, video streaming is the most popular as video traffic is predicted to represent as much as 82% of all consumer Internet traffic by 2021 from 73% in 2016 [1]. Such a massive amount of the multimedia traffic in combination with the quality demanding users requires Network Operators (Internet Service Providers (ISPs) and Mobile Network Operators (MNO)), to include Quality of Experience (QoE) in the network management [2], [3]. Indeed, QoE represents the quality perceived by the user with the inclusion of multiple influencing factors, such as: human, context, business, application, and network [4]. An improvement of the QoE has a positive impact on the user churn and then on the revenue of the provider. However, the provision of higher QoE to the users requires network operators to incorporate novel QoE- aware network management approaches [5]–[8]. As this paper focuses on the 5G network, in the following, we refer to the MNO as the operator that owns and manages the network.

Concerning video streaming services, most of the OTT video providers currently rely on HTTP adaptive streaming (HAS). A standard HAS is the Dynamic Adaptive Streaming Over HTTP (DASH) issued by MPEG [9]. With HAS, client video players estimate the available channel bandwidth and adapt the video bitrate accordingly to avoid playback interruptions due to buffer underruns [10]. However, the OTT provider has no control over the end-to-end network, which may introduce excessive throughput fluctuations resulting in frequent quality switches, which degrade the quality perceived by the end user [11]. Nonetheless, recent approaches consider network-assisted streaming strategies, such as the one pro- posed by the Server And Network Assisted DASH (SAND DASH) [12]. The SAND standard enables an opportunity for MNO-OTT collaborative service management in video stream- ing. SAND interfaces allow signalling between the client and network control functions through which the network con- troller may get QoE/client player stats related information and may also assist the client in selecting optimal bitrates based on available network resources [12]. Furthermore, the SAND may enable the MNO to activate network control mechanisms (e.g., flow prioritization and network slicing) utilizing QoE/client player stats related information. For example, the study in [13] utilizes the SAND in combination with the Software-Defined Network (SDN) controller to provide better QoE. Thus, the SAND may lead to the deployment of collaborative QoE management approaches.

Zero-rating is a common practice in the multimedia communication industry these days, especially the zero-rated data rate version. In the zero-rated data rate approach, the MNOs limit the data rate of the traffic from the collaborating OTTs to the users (who accepted this service model) with the advantage for the latter that the relevant traffic is not counted when considering the maximum amount of traffic allowed per month (per contract). In this scenario, the MNO utilizes the

TABLE I
ACRONYMS USED IN THE ARTICLE.

Abbreviation	Description	Abbreviation	Description	Abbreviation	Description
ABR	Adaptive BitRate	MNO	Mobile Network Operator	(R)AN	(Radio) Access Network
AF	Application Function	MPD	Media Presentation Description	RRUs	Remote Radio Units
AMF	Access and Mobility Management Function	NEFs	Network Element Functions	SAND	Server And Network Assisted DASH
AUSF	Authentication Server Function	NFV	Network Function Virtualization	SD	Standard Definition
CLV	Customer LifeTime Value	NQA	Network-aware Quality Assistant	SDN	Software Defined Networks
CoS	Class of Service	NSSF	Network Slice Selection Function	SEBRA	SAND-Enabled centralized Bitrate and Resource Allocation
DANE	DASH-aware Network Element	OFDM	Orthogonal Frequency-Division Multiplexing	SLAs	Service Level Agreements
DASH	Dynamic Adaptive Streaming Over HTTP	OTT	Over-The-Top	SLOs	Service Level Objectives
DN	Data Network	PCF	Policy Control Function	SMF	Session Management Function
DPI	Dots per Inch	PR	Premium users	ST	Standard users
ELAs	Experience Level Agreements	PRBs	Physical Resource Blocks	UDM	Unified Data Management
FHD	Full High Definition	QAM	Quadrature Amplitude Modulation	UE	User Equipment
HD	High Definition	QMS	QoE Metric Server	UPF	User Plane Function
ISP	Internet Service Provider	QoE	Quality of Experience		
MEC	Mobile Edge Computing	QoS	Quality of Service		

MPEG DASH SAND standard for the DASH video streaming applications [12]. The significant advantage of the zero-rated data rate approach is that it allows the MNOs to reduce the video traffic load in the network, especially the radio network, which is often congested mostly due to the ever-increasing bandwidth demand from the users of the video streaming applications. A significant example of its application is the T-Mobile Binge On, where the MNO limits the maximum data rate (throughput) for the users as required by the low-resolution video representation [14]. Notwithstanding the mentioned benefits of the proposed approach, it is unable to control the QoE provided to the final users as it does not differentiate the provided resources with respect to the device, content resolution, and Class of Service (CoS) that characterizes each user.

Based on the mentioned limitations of the zero-rated data rate approach, in this paper, we propose the zero-rated QoE approach, which is based on the same underlined principle but instead of limiting the data-rate, it limits the allocated resources by considering the provided QoE level. The contributions of this work are the following:

- We introduce the novel concept of zero-rated QoE approach for video streaming services, which has the same benefits of the existing zero-rated approach while improving the overall quality provided to the users when using the same resources.
- We define how the proposed MNO-OTT collaboration can be introduced in the 5G network architecture by describing the functionalities of the novel blocks that need to be introduced and the flow of data that characterizes their interactions.
- The radio resource management algorithm is described to implement the proposed approach, which considers the CoS, QoE and user-device type for the resource management.
- Extensive simulation-based analyses are provided, which show significant improvements of the proposed algorithm in terms of delivered QoE and QoE-fairness.

The paper is structured as follows. Section II discusses related work. Section III proposes the reference architecture for the collaborative QoE management of the OTT video streaming application in 5G networks. Section IV presents the proposed zero-rated QoE approach and the radio resource allocation algorithm for OTT video streaming delivery. Section V provides the experiments setup and results. Finally, Section VI concludes the paper. Table I lists the acronyms used in this article.

II. Related Works

This section provides the insight into the state-of-the-art related to the present work. The section is further divided as follows: Section II-A provides the details of the 5G architecture considering the latest standardization activities and state-of-the-art works done for the QoE management of the multimedia services; section II-B investigates the state-of-the-art towards collaborative service management that takes into account the QoE. Table II represents the summary of the related state-of-the-art works from the literature.

A. 5G Networks and QoE Management

The cloud-native architecture of the 5G networks provides an opportunity for the data-driven personalized QoE management [15]. Though numerous efforts in the state-of-the-art can be found related to QoE management in 5G paradigm [25], [26], we would like to highlight most relevant studies from the literature with regards to data-driven approaches for the QoE management in the 5G networks. The work in [15] proposed a data-driven 5G architecture for the personalized QoE management utilizing mobile agent-based QoS/QoE monitoring for the user preferences, context factor, and user's experience. The data mining component utilized the collected data from the mobile agent for the QoE prediction. The network elements

TABLE II
SUMMARY OF RELATED STATE-OF-THE-ART WORKS.

State-of-the-art work	Contribution
5G Networks and QoE Management	
Wang *et al.* [15]	Data-driven 5G architecture for QoE management Mobile agent based QoS/QoE monitoring
Agyapong *et al.* [16]	Design consideration of the cloud-native 5G architecture The concept of Network as a service to OTT provider
Gramaglia *et al.* [17]	Architecture for QoE management in 5G networks QoE monitoring at user-terminal
Jiang *et al.* [18]	QoE-based admission control QoE-aware prioritization
Dutta *et al.* [19]	5G-network cloud management using QoE Cost minimization of network cloud
Ge *et al.* [20]	QoE-driven caching and network management for video streaming services
Collaborative QoE Management of Video Streaming	
Cofano *et al.* [21]	Strategies for network-assisted video streaming services
Khorov *et al.* [22]	SAND-Enabled centralized Bitrate and Resource Allocation in wireless networks
Varela *et al.* [23]	ELAs with the customers to deliver the guaranteed QoE
Floris *et al.* [7]	Architecture for Collaborative QoE management by OTT-ISP Three OTT-ISP collaboration strategies Inclusion of business model in the service delivery
Ahmad *et al.* [24]	Architecture for the QoE-driven OTT-ISP collaboration CLV based service management strategy

utilize the predicted QoE for network-wide QoE management. The proposed work contributes towards QoE monitoring by the inclusion of the QoE related influencing factors such as user preferences and context factor. However, the work does not address the QoE management procedure and the business model influence on QoE (which remains more important according to industrial perspectives [27]).

The study in [16], provided a vision of the design consideration of the cloud-native 5G architecture where the network architecture composed of two logical layers: network cloud and radio network. The network cloud includes the user and control plane entities which perform higher layer functionalities related to resource allocation and traffic engineering while the radio network performs minimum L1/L2 functions. An important concept is introduced where MNO can offer "Network as a Service" to OTT providers through northbound APIs to offer better QoE delivery to their customers. In [17], the work proposes an architecture for the QoE management by extracting the QoE related information from the user terminal for QoE monitoring. Moreover, the work introduces an essential concept of distributed service management by multiple Software-Defined Network (SDN) controllers for inter-slice and intra-slice resources where QoE monitoring is done at the intra-slice level through the SDN controller while QoE management at intra-slice is performed by exploiting QoE/QoS mapping. However, the work does not provide any details regarding the QoE-aware collaborative management of OTT services in 5G architecture. Furthermore, the work conducted in [18], proposes QoE-based admission control algorithm for 5G networks at intra-slice and inter-slice levels by utilizing the network slicing function offered by the Network Function Virtualization (NFV). In addition to the QoE-aware admission control algorithm, the work in [18] also proposes QoE-aware prioritization on both intra-slice and inter-slice level. The major limitation of this work is as follows: main influencing factors on QoE are ignored, such as business, pricing, context. This work does not provide insight on how the QoE monitoring can be performed in 5G architecture and how collaborative QoE management of OTT service can be performed in the era of end-to-end encryption. Similarly, the study in [19] proposes the management and scaling of network cloud resources from the delivered QoE to the users to lower down the cost of the cloud infrastructure. However, the work does not highlight how the QoE related information is extracted neither what are the interfaces required by the MNO for the collaborative QoE management of the OTT services. Likewise, the study in [20] proposes QoE driven caching and network adaptation for the OTT video streaming while ignoring the information exchange between the OTT and MNO.

In summary, the state-of-the-art works mentioned above, propose important concepts for the QoE management of OTT video streaming in 5G however, following key issues has not been addressed: 1) Inclusion of the business and pricing related influence factor on QoE. 2) How the collaborative QoE management of the OTT services can be performed? 3) What information is needed to be extracted to predict QoE? 4) What are the information exchange points/interfaces for the QoE monitoring of the OTT services? These questions are addressed in this article.

B. Collaborative QoE Management of Video Streaming Application

Most of the studies proposed for QoE management of video services are based on HAS, with which client video players can adaptively select the quality of downloaded video by the estimated current channel capacity. Besides the DASH standard [9], many bitrate adaptation algorithms aimed at improving the QoE have been proposed in the literature [10] [28]. The major drawbacks of these approaches are: i) clients have no information about the channel capacity at the beginning of a video flow; ii) clients may incorrectly estimate available channel capacity; iii) network actions are not considered [22]. Network-assisted streaming strategies have been defined to address these issues. In [21], the performance of three classes of network-assisted strategies are compared, which includes bandwidth reservation per video flow, bandwidth adaptation driven by video bitrate and a combination of both. It results that all the considered approaches provide remarkable improvement in obtained video quality fairness when compared to the case in which any network-assisted strategy is used. In particular, the second approach provides the best results. In 2017 MPEG published an extension to the DASH standard called SAND [12], with which DASH-clients and DASH-servers can exchange control messages with intermediate network nodes called DASH-Aware Network Elements (DANEs). DANEs are logical entities which are set up inside the network to manage resources at the bottleneck links. However, the impact of the business strategies, client-side device, and extraction of the QoE-aware information for

the OTT video streaming services are not considered in the presented QoE management approach. Furthermore, the work is not focused on 5G networks. In [22], a SAND-Enabled centralized Bitrate and Resource Allocation (SEBRA) algorithm is proposed for network-assisted video streaming in wireless networks aiming to reduce initial buffering delay and to avoid stalls by accelerating video download for Wi-Fi stations with almost empty buffers. The results have shown that SEBRA can significantly improve QoE in comparison with the state-of-the-art DASH management algorithms, especially for short video flows. Nevertheless, important QoE influencing factors such as business, device type, and context are missing in this work. Moreover, the work is not focused on 5G networks.

The work in [23] proposes that the QoE-aware service delivery cannot be done with the normal Service Level Agreements (SLAs), but it requires Experience Level Agreements (ELAs) with the customers to deliver the guaranteed QoE. Thus, the ELAs may require the change in the Service Level Objectives (SLOs) among OTTs and MNOs. In our previous work [7], the study discusses the collaborative QoE management by the OTT and ISP. In this work, three different collaborative QoE management strategies for the OTT applications are proposed where the impact of the collaborative business strategies and pricing on the delivered QoE, user churn and profitability of the approach is studied. The proposed approach used the cloud information exchange as an interface between the OTTs and ISPs to exchange QoE-related information among each other, which drives the collaborative QoE-management of OTT services. However, the work is not focused on 5G networks. Similarly, our work in [24] considers the collaborative QoE management of the OTT video streaming application in the ISP network by the management of the surrogate servers in the ISP network. The customers are classified into two categories based on the Customer Lifetime Value (CLV). The ISP manages the surrogate servers in the different location zones to deliver better quality to the users with higher CLV. Nonetheless, the proposed approach is not specifically designed for 5G networks.

Network Neutrality (also called Net Neutrality) stands an important concept when it comes to the collaborative service management of the OTT services. Although there is no standard definition of Net Neutrality yet, however according to Net Neutrality, ISP should not discriminate/block any content or application [29]. Thus based on the Net Neutrality principle, ISP should treat OTT traffic as best effort however low-level Net neutrality violations such as traffic prioritization and load management can be accepted [29]. Furthermore, the work in [30] classifies Network Neutrality as a potential threat to future innovation and technology in network management as it eliminates ISP's perspectives to invest in the network infrastructure. The ongoing industrial collaborations such as zero-rated data rate approach, e.g., T-Mobile Binge On where ISP limits the data rate of the collaborating OTT provider can be considered as a low-level violation of the Net Neutrality [14]. Similarly, the collaboration schemes such as Google Global Cache[1] and Netflix Open Connect[2] allows ISP's to collaborate with the Google and Netflix where ISP hosts the surrogate servers from these OTTs. The surrogate servers act as a cache and reduce unnecessary traffic in the core network. However, we believe that such kind of collaboration can be categorized as a low-level violation of the Net Neutrality principle, as this approach reduces the latency only when accessing the specific content hosted in the ISP premise [29]. Furthermore, there are different global views on the Net Neutrality i.e., Net Neutrality has been repealed in United State since June, 2018[3] while European Union still tries to enforce it. 5G is also considered to be a threat to Net Neutrality as application-specific network slicing in 5G may treat the Internet traffic differently[4]. Therefore, a better definition of Network Neutrality is indeed needed.

III. 5G ARCHITECTURE FOR COLLABORATIVE QOE MANAGEMENT

This section describes the proposed reference architecture for the QoE-driven collaborative service management for video streaming applications in the 5G network. In the proposed reference architecture, QoE related information is acquired from the end-user probe installed in the terminal where the OTT application is running, as inspired by our previous works [31]–[33]. The proposed approach relies on the following roles assumed by the different entities involved in the multimedia service delivery: 1) *User*, who allows for information retrieval from her terminal (UE - User Equipment) through the installed probe where the OTT video streaming application is running, taking into account her privacy preferences; 2) *OTT*, which provides RESTful APIs to the ISP to get information from the UE probe; 3) *MNO*, which extracts QoE influence factors data via RESTful APIs provided by the OTT and makes use of this for QoE-driven network management. Fig. 1 represents the proposed reference architecture. Herein, the overall architecture follows the standard 3GPP specifications with all the components involved keeping the standard defined functionalities [34]. With this basis, we have added the components that are needed for the implementation of the proposed strategy, which are highlighted with blocks with red dashed borders in the figure. Additionally, Fig. 2 shows the interaction among the proposed blocks.

A. User Equipment and OTT Components

The UE is crucial for the acquisition of key information from the installed probes. It is important to highlight that this has to happen with the consensus of the user, which has to be informed about the type of data retrieved and how this is treated. In any case, only anonymous data is retrieved.

For the proposed solution, the OTT application at the UE contains three main blocks: 1) the DASH Client, which

[1]https://peering.google.com

[2]https://openconnect.netflix.com

[3]https://money.cnn.com/2018/06/11/technology/net-neutrality-repeal-explained/index.html

[4]https://thenextweb.com/eu/2019/02/28/5g-is-a-threat-to-europes-absolute-net-neutrality/

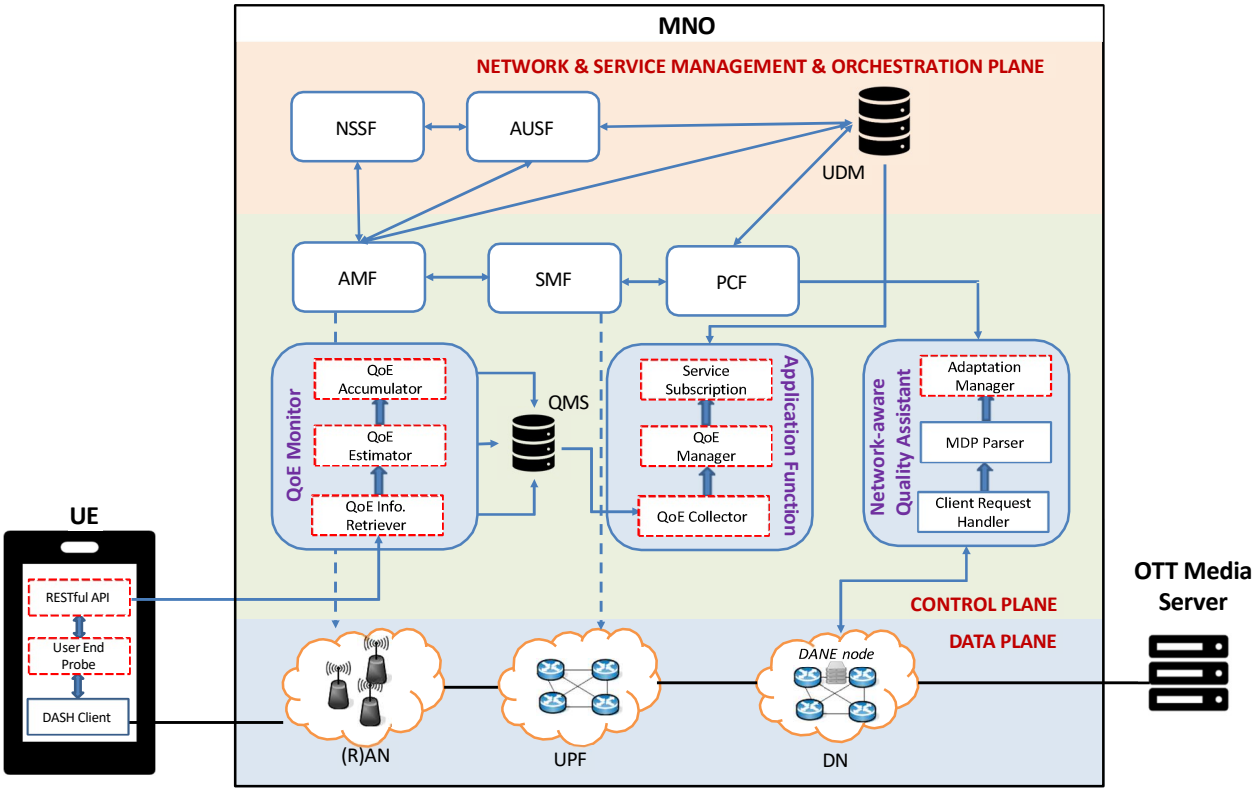

Fig. 1. Proposed architecture which relies on the 3GPP architecture [34]: the blocks with red dashed edge are those introduced for the implementation of the proposed solution.

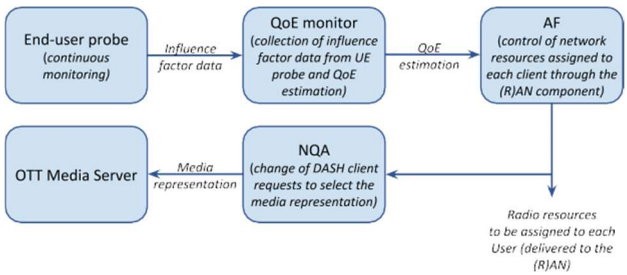

Fig. 2. Diagram showing the interaction among the proposed architectural components.

supports the communication with the Network-aware Quality Assistant (NQA) module in the control plane through web sockets following the MPEG DASH SAND standard [12]. All the client requests for the video playback are handled by the NQA, which is discussed in detail in Section III-B2; 2) the UE probe, which collects QoE influence factors data from the user terminal and; 3) the RESTful API, which provides the interface for the retrieval of QoE collected data, which is handled by the QoE Information Retriever as highlighted in Fig. 2. The OTT Media Server provides the DASH server functionalities and the content without any specific additional functionalities introduced for the proposed solution.

B. MNO Components

1) Data Plane: The data plane of the proposed architecture is composed of four components: UE, (Radio) Access Network (R)AN, User Plane Function (UPF) and Data Network (DN). The (R)AN is the radio network part of the 5G network that is composed of the radio base stations and the Remote Radio Units (RRUs). The (R)AN connects the UE of the network slice with the MNO's core cloud network (UPF). The UPF is the MNO's core cloud network which is composed of the Mobile Edge Computing (MEC) resources and network core of the network slice and connects the network slice with the external Internet which is represented by the DN. In the DN, an open switch is programmed to act as a DANE component as controlled by the NQA. Accordingly, it replaces the DASH packets generated by the UE and transmitted at the data plane to control the streaming session rate.

2) Contol Plane: The control plane is composed of the Network Element Functions (NEFs), which control the network elements in the data plane within the dedicated slice. The Access and Mobility Management Function (AMF) is responsible for the management of the (R)AN in the network slice, while the Session Management Function (SMF) performs the management of the core cloud network (UPF) in the network slice. The Policy Control Function (PCF) enforces the network policies within the network slice by providing the policy rules to the control plane NEFs. In the proposed architecture, the PCF enforces the network policy on the basis of the outcome of the Application Function (AF), as it is explained in the

following.

According to the 3GPP specification TS 23.501 [34] for the 5G network, for the radio resource allocation, the UE requests session establishment from the AMF. After the authentication and based on PCF policies, the AMF sends the resource setup request to the (R)AN (gNode B) to allocat radio resources to the UE. The (R)AN then sends the radio resource control reconfiguration to the UE, which is followed by the UE radio resource configuration complete response from the UE to the (R)AN. After the session resource setup response from the (R)AN to the AMF, the radio resources are allocated to the UE, and the session is established.

The QoE Monitor performs the QoE monitoring and measurement of the video streaming applications within the network slice. It is composed of three modules: 1) the QoE Info Retriever, which retrieves the QoE related information from the UE through HTTP GET requests; 2) the QoE Estimator, which estimates the QoE by utilizing the information collected by the QoE Info Retriever module; 3) the QoE Accumulator, which computes the average QoE predicted during the client sessions over time. All the information from the QoE Monitor is stored in the QoE Metric Server (QMS), and this information is exploited by the AF for network management. It is important to highlight that the retrieval of the information about the streaming session is possible only with the collaboration between the two providers. Indeed, due to the encryption of the data from and to the media server, the MNOs will not be able to get influence factors data from the UE through the OTT APIs.

The AF is responsible for the implementation of the strategy proposed in this paper in the QoE Manager component. The latter relies on the data extracted from the QMS server and the information about the user profiles obtained by the Service Subscription component. It then computes the optimal radio resources to be assigned to each user and the corresponding streaming rate. The outcome is fed into the PCF.

The NQA module is responsible for assisting the DASH client-side player in adapting the video quality resolutions and bitrates following the MPEG DASH SAND specifications. It consists of the following three modules: 1) the Client Request Handler, which listens to the client request and passes it to the subsequent high-level modules for further processing (this happens when the DASH client at UE initiates the video playback request of the OTT video streaming service). Moreover, this module delivers the outcome of the adaptation manager to the DASH client; 2) the MPD Parser, which parses only representation information of the Media Presentation Description (MPD) file of the requested video from the DASH player. It includes attributes of video content to be played, such as video quality resolutions, bitrate, number and length of the segments. Note that only the above-mentioned representation details of the MPD files are parsed without content-related details. The MPD parser analyzes the MPD file and sends the outcome to the Adaptation Manager module; 3) the Adaptation Manager, which overwrites the representation details of the MPD file and assists the client in adapting video quality/bitrate on the basis of the feedback provided by the Policy Control Function (PCF) following the outcome of the QoE Manager. The adaptation manager modifies the client request by selecting the video resolution and the bitrate accordingly. Note that only the Adaptation Manager in the NQA is considered to provide novel functionalities and is then highlighted accordingly in Fig. 2, as the other two components are already present in a SAND solution.

3) Network & Service Management & Orchestration Plane: This layer is composed of three NEFs: 1) the Network Slice Selection Function; 2) the Authentication Server Function (AUSF); 3) the Unified Data Management (UDM). During the connection establishment phase, the Network Slice Selection Function (NSSF) selects the network slice and connects the user to a particular slice based on the user's subscription and requested service after the confirmation from the AUSF. The AUSF authenticates the users by relying on the data stored in the UDM.

C. Discussion on Accuracy and Scalability

The implementation of the proposed solution can leverage on the NFV and MEC technology enablers characterizing the 5G networks. The NFV functions can address the scalability of the proposed architecture by the placement of the multiple distributed virtualized NEFs in the virtual network infrastructure. These allow for allocating the appropriate resources according to the varying computational requirement of the proposed solution that depends on the network load. To further address the scalability, the QoE monitor in the proposed architecture can be located close to the user probes as virtual NEF at the network edge using MEC technologies. The placement of the QoE monitor using the MEC resources allows for estimating the QoE locally (close to the user) and deciding whether there is a need for activating control actions. This allows for reducing the traffic from the user probe to a central NQA component; only control signals, when particular conditions are reached, are then sent. In the proposed architecture, the accuracy of the monitored QoE highly depends on the monitoring frequency of the UE probe: the higher the frequency, the higher the accuracy, as also discussed in our previous work [35]. However, a trade-off between accuracy, the data generated by the monitoring probe and latency in control plane operations is required.

IV. Zero-Rated QoE Approach

Exploiting the architecture that we have described in the previous section, herein we illustrate a new collaborative approach which we call zero-rated QoE, which is inspired by the zero-rated data rate approach. Table III presents the variables used in the article.

As already discussed in the Introduction Section, the zero-rated data rate approach is motivated by the issue for the MNOs to face an ever increasing bandwidth demand from the users, mostly due to the video streaming applications, which often brings to radio network congestion events. With such an approach the MNO can bring the users to accept to limit her download bandwidth by not counting the relevant traffic in her contract monthly data limit. In this way, the MNOs are expected to avoid or deeply reduce the radio network

congestion events. This is an alternative to other network resource management approaches, where the network operator (in case of network resource limitations) reduces the user throughput without the consent of the user but whenever the congestion is reached. Several principles have been proposed in the past, which (when QoE is considered) may consider the objective of maximizing the average user estimated QoE, the percentage of the users with a QoE lower than a given threshold or similar. Several techniques are reviewed in [2] and [8]. The zero-rated data rate approach has the advantage of reaching an agreement with the user so that a reduction in quality comes with practical benefit (which can be translated into an economic advantage) for the user and it is also in-line with the expectations of the users who know of this bandwidth restriction. These two aspects bring to an improvement of the user QoE when compared to an approach where bandwidth is throttled without user awareness.

In this work, we leverage on the benefits of this approach and we address the following primary limitations of the zero-rated data rate approach: 1) *Video characteristics* are not considered in the resource allocation, i.e., different content types require different levels of network resources [36]; 2) The *end device resolution* is also not taken into account in the service optimization, whereas it is well-known that when low resolution video is streamed, the user with a high-resolution device perceives a lower quality with respect to the one using a low-resolution device [37]; 3) *QoE* is not considered neither as a part of the service plan nor in the resource allocation for the service optimization ; 4) the impact of *service subscription plan* and Class of Service (CoS) has also been ignored in the resource distribution, whereas users who pay more expect better quality [38].

Therefore, we propose the zero-rated QoE approach which relies on the same primary principle as the zero-rated approach but instead of the zero-rated data rate, zero-rated QoE level is provided to the users based on the QoE model for video streaming services. As this approach requires the MNO to be able to estimate the QoE, there is the need for it to access the data of relevant influence factors. In the proposed scenario this is possible only because a collaboration exists between the MNO and the OTT, which allows for overcoming the limitation imposed by the encryption of the data to and from the media server. Moreover, the proposed approach improves the major limitation of the zero-rated data rate approach by considering CoS, video characteristics and device type in the resource allocation algorithm.

A. Resource Allocation Algorithm

We suppose that the zero-rated QoE based video streaming service is offered in total K CoS, where $k = 1, 2, ..., K$ indexes the CoSs from the highest to the lowest class priority. The different CoSs are considered in the approach because users may have different service subscription plans with the MNO. Also, the users who pay more expect more [39]. Therefore, different zero-rated QoE levels are considered for different CoSs, where QoE_k represents the zero-rate QoE level on ITU-T ACR scale $(1-5)$ for the k-th CoS. The

TABLE III
VARIABLES USED IN THE ARTICLE.

Variable	Description
$k = 1, 2, 3, ..., K$	Index of CoS
$n = 1, 2, 3, ..., N$	Index of video representations
β	Total available capacity (in PRBs)
$r = 1, 2, 3, ..., R$	Index of device resolution
$i = 1, 2, 3, ..., I$	Index of users
QoE_k	Target QoE level for users of CoS k in the proposed approach
T	Maximum estimated throughput
BR	Bitrate of the video
BR^*	Selected optimal bitrate
α	Weighting factors to maximum bitrate
BW	Estimated bandwidth
SNR	Signal to Noise Ratio
F	Fairness index
QoE_{ST}	Target QoE level for standard users in the proposed approach
QoE_{PR}	Target QoE level for premium users in the proposed approach
T_{ZR}	Target throughput in the zero-rated data rate approach

conventional DASH player adaptation algorithm selects the video playback bitrate based on the available throughput [21] (here we consider PANDA as a conventional adaptive bitrate algorithm which is proposed in [40]), i.e., playback bitrate closer to the corresponding available throughput is selected. Considering this, we assume that the videos being played have $n = 1, 2, 3, ..., N$ representations for a given resolution in different bitrates, where N is the highest representation with the maximum bitrates available (highest layer).

Mostly, in the cellular networks, the quality degradation and resource limitation happen at the last mile of access network [41]. Therefore we consider that the MNO has fixed capacity RAN resources with reference to the Physical Resource Blocks (PRBs). We assume that the total available capacity of RAN in terms of PRBs is β and that there is a total of $I_{r,k}$ users belonging to the k-th CoS with a device with resolution r ($r = 1, 2, 3, ..., R$ indexes the device resolution in Dots per Inch (DPI), where R is the highest device resolution). Then, the optimal bitrate $BR^*_{i_{r,k}}$ is selected for the i-th user with the r device resolution belonging to the k-th CoS on the basis of the objective function in Eq. (1):

$$\min_{BR_{i_{r,k}}} \; | QoE_{i_{r,k}} - QoE_k | \tag{1}$$

$$s.t.\ 0 \le BR_{i_{r,k}} \le \max(BR_{i_{r,k}})$$

where $BR_{i_{r,k}}$ are the bitrates of the r resolution of the video being played by the i-th user from k-th CoS while $\max(BR_{i_{r,k}})$ is maximum bitrate available for r resolution of that video. $QoE_{i_{r,k}}$ is the estimated quality level at $BR_{i_{r,k}}$ bitrate and QoE_k is the target quality (parameter set in the proposed approach). Then, the maximum estimated throughput $T_{i_{r,k}}$ to be delivered to the i-th user belonging to the k-th CoS with r device resolution can be computed according to Eq. (2):

$$T_{i_{r,k}} = \alpha \cdot BR^*_{i_{r,k}} + (1-\alpha)\frac{\sum_{I_{r,k}} BR^*_{i_{r,k}}}{I_{r,k}} \tag{2}$$

where α is the weighing factor to the optimal selected bitrate. In order to allocate PRBs to the user, the bandwidth (in Hz) to be assigned to the i-th user is computed by the Shannon capacity formula in Eq. (3):

$$BW_{i_{r,k}} = \frac{T_{i_{r,k}}}{log_2(1 + SNR_{i_{r,k}})} \quad (3)$$

where $SNR_{i_{r,k}}$ is the minimum Signal-to-Noise Ratio (on linear scale) of the transmission channel of the i-th user. According to the 3GPP specification TS.38.211 [42], 1 PRB contains 12 sub carriers, while each sub carrier is characterized by a 15 KHz bandwidth. Therefore $1\ PRB = 12 \cdot 15KHz = 180KHz$. Thus, the PRBs allocated to the i-th user can be computed by Eq. (4):

$$\beta_{i_{r,k}} = \frac{BW_{i_{r,k}}}{180 \cdot 10^3} \quad s.t.\ \beta \geq 0 \quad (4)$$

In the proposed resource allocation algorithm, the resources are allocated to the users in the session initiation phase in terms of PRBs. The priority in resource allocation is given to users with the higher CoS and lower resource requirements, i.e., the user with the lowest device resolution and highest CoS (who pay more price) are given highest priority and so on. If physical resources are utilized to the maximum limit, admission control is triggered to stop further degradation of the quality. Algorithm 1 describes the steps to be followed by the proposed resource allocation algorithm.

As to the zero-rated data rate approach, the allocated resources are calculated by assigning the same target throughput to all the users and set equal to T_{ZR}. This value is then used to compute the bandwidth following Eq. (3) and then from this to compute the allocated PRBs following Eq. (4).

B. Resource Allocation Algorithm mapping to Proposed Architecture

During the video session initiation phase, the NQA retrieves the MPD file of the requested video through the RESTful API HTTP GET response from the UE probe. The retrieved MPD file is then parsed by the relevant parser block, which extracts information regarding media representations resolutions and respective bitrates for the estimation of the throughput to be delivered to the user. The MPD parser sends the extracted resolutions and bitrates of the media representations to the QoE estimation block in the QoE Monitor. Moreover, the QoE monitor block collects the information regarding the user's device resolution from the UE via the UE probe. Furthermore, the QoE estimation block accesses service plan subscription and CoS information from the UDM and performs the selection of the video bitrates considering the user's device resolution and CoS. The bitrate selection is performed by the QoE estimation block using Eq. (1). The feedback from the QoE Estimator block is provided to the resource allocation block in the AF, which computes the estimated throughput to be delivered, required bandwidth and PRBs to be allocated to the user using Eq. (2)-(4). In the case of resource availability, the AF allocates the resources to the user by the AMF through the PCF feedback. After the session initiation, the adaptation manager assists the client player to play video segments of the video representation selected by the QoE Estimator block.

Algorithm 1: Resource Allocation Algorithm

Input : k; K; α; β; r; R; I_k; $\forall nBR_{i_{r,k}}$; $SNR_{i_{r,k}}$; QoE_k

Output: Number of PRBs allocated to $i_{r,k}$ ($\beta_{i_{r,k}}$)

```
1  foreach k → K do
2      foreach r → R do
3          foreach i_{r,k} → I_{r,k} do
4              selects BR*_{i_{r,k}} using Eq. (1)
5              computes T_{i_{r,k}} with Eq. (2)
6              computes BW_{i_{r,k}} with Eq. (3)
7              computes β_{i_{r,k}} with Eq. (4)
8              if β − β_{i_{r,k}} < 0 then
9                  allocate computed β_{i_{r,k}} in Step 7
10                 allocate computed BW_{i_{r,k}} in Step 6
11                 β = β − β_{i_{r,k}}
12             else
13                 BW_{i_{r,k}} = 0
14                 β_{i_{r,k}} = 0
15             end
16             return β_{i_{r,k}}
17         end
18     end
19 end
```

V. Simulations

The simulations are conducted to investigate the effectiveness of the proposed zero-rated QoE approach in comparison with the zero-rated data rate approach on the basis of: 1) QoE delivered to the user; 2) QoE-Fairness among the users and; 3) Network resource utilization. Section V-A provides the details of the evaluation metrics considered in the comparative analysis; Section V-B describes the details of the simulation setup; and the Section V-C analyzes the results of the conducted simulations.

A. Evaluation Metrics

1) QoE Model: For QoE prediction, we consider the ITU-T P.1203 standard [43], which is a parametric bit-stream based quality assessment/prediction model for HTTP adaptive video streaming. The simulations use ITU-T P.1203 in Mode 0 due to the fact that information related to device resolution, coding resolution representation bitrates and framerate are available to the MNOs in zero-rated QoE approach for the resource allocation. The ITU-T P.1203 model output $O.22$ (video coding quality per output sampling interval) is used to allocate radio resources in the zero-rated QoE approach. The evaluation of the results for both approaches is performed in terms of predicted QoE computed using the $O.46$ (final media session quality score in the ITU-T ACR scale ($1 - 5$)). For the simulations, the ITU-T P.1203 is implemented in MATLAB following the open-source Python language based implementation of the ITU-T P.1203 [44], [45].

TABLE IV
SIMULATION SETTINGS.

Parameters	Settings
Number of runs	100
Random number generator	Mersenne Twister
Random number seed	0
CoS and Population	
Number of CoS	2 (PR, ST)
Population size	60
Device types	3 (FHD, HD, SD)
Device resolutions	1920x1080 (FHD), 1280x720 (HD), 854x480 (SD)
Zero-rated QoE Approach	
QoE_{ST}	1-4 ITU-T ACR scale
QoE_{PR}	$QoE_{ST} + \delta QoE$
δQoE	0.5
α	0.75
Zero-rated Data rate Approach	
T_{ZR}	500-1500 (Kbps)
Videos and Network Settings	
Videos content types	Animation, documentary and movie
Videos play length	112 sec
Videos resolutions	1920x1080, 1280x720, 854x480, 480x360
Videos Frame rate	24 fps
Total capacity of network (β)	100 PRBs
OFDM Modulations	16 QAM, 64 QAM
SNR	12-20 dBs

2) QoE-Fairness Metric: In order to provide the comparison of the proposed approach with zero-rated data rate approach, the QoE fairness index proposed in [46] can be mathematically represented as:

$$F = 1 - \frac{\sigma}{\sigma_{max}}, \tag{5}$$

where σ is the standard deviation of the delivered QoE level while $\sigma_{max} = H - L$ is the maximum standard deviation in QoE level perceived by the users whereas H and L are the highest and lowest QoE levels. Considering the QoE level on ITU-T ACR scale (1-5), $H = 5$ and $L = 1$. Thus, the QoE fairness index among the $k - th$ CoS users with r resolution device is computed by the Eq. 6:

$$F_{r,k} = 1 - \frac{\sigma_{r,k}}{2} = 1 - \frac{1}{2}\sqrt{\frac{\sum_{i_{r,k}=1}^{I_{r,k}} (QoE_{i_{r,k}} - \overline{QoE_{r,k}})^2}{I_{r,k} - 1}} \tag{6}$$

where $I_{r,k}$ is the total number of users with r device resolution belonging to k-th CoS while $QoE_{i_{r,k}}$ and $QoE_{r,k}$ are the QoE level delivered to the i-th user and average QoE delivered to all users with r device resolution in the k-th CoS. Similarly, the fairness delivered among the users belonging to k-th CoS can be calculated using Eq.(7):

$$F_k = 1 - \frac{1}{2}\sqrt{\frac{\sum_{r=1}^{R}\sum_{i_{r,k}=1}^{I_{r,k}} (QoE_{i_{r,k}} - \overline{QoE_{r,k}})^2}{I_{r,k} - 1}} \tag{7}$$

B. Simulation Setup

Two scenarios for the service delivery are considered: 1) Zero-rated QoE approach and; 2) Zero-rated data rate approach. We performed Monte Carlo simulations on the MATLAB platform where experiments are repeated 100 times. In order to achieve repeatability of the results, MATLAB random number generator is set to the default settings which uses Mersenne Twister as random number generator with the seed 0. The proposed zero-rated QoE approach performs the resources allocation based on Algorithm 1 proposed in the Section IV-A. Table IV lists the parameters settings of the experiments.

1) CoS and Population: We consider two CoS of the service subscription where users are distributed based on the willingness to pay: 1) Premium (PR) subscription and; 2) Standard (ST) subscription. The subscription price for the PR service P_{PR} is higher than the ST service price P_{ST}. We set $P_{PR} = 0.6$ and $P_{ST} = 0.3$ based on our previous study in [38]. In the considered scenario, the users are distributed among different CoS randomly on the basis of willingness to pay (uniformly distributed between $0 - 1$), i.e. if the user's willingness to pay is higher than P_{PR}, the user is assigned to the PR CoS otherwise if the willingness to pay is greater than P_{ST} but less than P_{PR}, the user is allocated ST CoS. The population size for both approaches is considered to be 60 users (fixed), which are uniformly distributed between two CoS based on the willingness to pay. Moreover, user devices with three different resolutions such as Full High Definition (FHD) (1920x1080), High Definition (HD) (1280x720) and Standard Definition (SD) (854x480) are considered. The device resolutions are uniformly distributed among the users. In the start of each simulation run, the willingness to pay and devices types are uniformly distributed among the users then the users are assigned CoS based on the willingness to pay as mentioned above.

2) Videos and Network Settings: In the simulations, we consider three different types of videos from the ITEC database [47]. The video sequences of Big Buck Bunny, Of Forest And Men and Valkaama with the content type animation, documentary and movie respectively. All the video sequences have video play length of 112 seconds. For all videos, the frame rate is 24 fps. The MPD files of the videos with 4 seconds segments duration are included in the MATLAB simulations by the extracting the media representation information through the Python program. Table V provides detailed information regarding the video representations considered in the simulations. For the Adaptive BitRate (ABR), PANDA client-side rate adaptation algorithm presented in [40] is selected which is considered as conventional ABR [48].

We consider the Orthogonal Frequency-Division Multiplexing (OFDM) for transmission, where 16 Quadrature Amplitude Modulation (16 QAM) and 64 QAM are considered for digital modulation. For the cellular network, the total capacity in terms of the PRBs is fixed to $\beta = 100$, while SNR is uniformly distributed among the users within the range of $12 - 20$ dBs. It should be noted that air gaps and path losses are also considered in this range of SNR values. The selection of

TABLE V
REPRESENTATIONS DETAILS OF THE VIDEOS IN SIMULATION

Video Representation Bitrates (Kbps)			
Resolutions	**Big Buck Bunny**	**Of Forest And Men**	**Valkaama**
480x360	177.437 255.865 378.355	172.627 242.301 298.523	172.453 215.003 330.491
854x480	509.091 577.751	505.903 574.104	754.258 930.297
1280x720	1008.699 1207.152 1473.801	983.335 1217.742 1422.377	1323.244 1716.694 1988.387
1920x1080	2409.742 3340.509 3936.261	2317.655 3522.543 4118.338	2708.994 3430.035 4003.428

TABLE VI
MEAN PREDICTED QOE DELIVERED W.R.T COS AND DEVICE TYPES ($QoE_{PR} = QoE_{ST} + \delta QoE$ WHERE $\delta QoE = 0.5$ AND $\alpha = 0.75$).

Zero-rated QoE approach						
QoE_{ST} (MOS)	Mean Predicted QoE (MOS)					
	PR FHD	PR HD	PR SD	ST FHD	ST HD	ST SD
2.0	2.5	3.1	3.2	2.4	2.7	3.5
2.2	2.7	3.0	3.2	2.5	2.8	3.5
2.4	2.7	3.3	3.2	2.5	3.0	3.5
2.6	3.1	3.3	3.5	2.7	3.0	3.5
2.8	3.1	3.3	3.6	2.7	3.2	3.5
3.0	3.5	3.6	3.9	3.0	3.4	3.6
3.2	3.7	3.9	3.9	3.2	3.4	3.7
3.4	4.0	4.0	4.0	3.1	3.5	3.7
3.6	**4.0**	**4.1**	**4.1**	**3.7**	**3.6**	**4.0**
3.8	4.2	4.1	4.1	4.0	4.0	4.0
4.0	4.2	4.1	4.1	4.0	4.0	4.1
Zero-rated data rate approach						
T_{ZR} (Kbps)	Mean Predicted QoE (MOS)					
	PR FHD	PR HD	PR SD	ST FHD	ST HD	ST SD
500	2.3	3.0	3.7	2.2	3.1	3.5
600	2.7	3.3	3.8	2.6	3.3	3.6
700	2.7	3.2	3.8	2.6	3.3	3.7
800	2.9	3.6	3.9	2.9	3.5	3.9
900	2.9	3.7	3.9	2.9	3.7	3.9
1000	2.9	3.6	3.8	3.0	3.6	3.8
1100	**3.3**	**3.7**	**3.9**	**3.1**	**3.8**	**3.8**
1200	3.3	3.7	3.8	3.1	3.8	3.8
1300	3.5	3.8	3.8	3.4	3.8	3.8
1400	3.5	3.8	3.9	3.4	3.8	3.9
1500	3.6	3.8	3.9	3.5	3.8	3.9

the SNR in this range is based on the 3GPP standard TS 36.101 [49], i.e., for SNR higher than 18 dBs, the 64 QAM modulation is selected, whereas when the SNR is in the range $5 < SNR < 18$ dBs, the 16 QAM modulation is adopted.

At the beginning of each simulation run, the user are randomly assigned SNR value in the aforementioned range, which is followed by the radio resource allocation. The video sequences to be played by the users are randomly assigned. All the user requests for the video sessions simultaneously after the radio resource allocation. The video sessions last until the length of videos (112 seconds). The QoE is predicted over the video session using ITU-T P.1203 model [43].

C. Results and Analysis

This section discusses the simulation results by providing the comparative analysis of the proposed zero-rated QoE approach (presented in Section IV with the zero-rated data rate approach.

Table VI provides the comparative analysis in terms of mean predicted QoE delivered to the users belonging to different CoS and device types. In the proposed approach, we vary the target zero-rate QoE level of ST users QoE_{ST} from $2-4$ (in the ITU-T ACR scale $(1-5)$) with the step size of 0.2; for the PR users the target zero-rated QoE has been set as follows: $QoE_{PR} = QoE_{ST} + \delta QoE$, where $\delta QoE = 0.5$ for all the experiments. Similarly, for the zero-rated data rate approach, the target throughput T_{ZR} is varied in the range $500-1500$ Kbps, with the step size of 100 Kbps. For all the experiments, α is set to 0.75. It can be noticed that the zero-rated QoE approach delivers mean QoE approximately equal to the QoE_{ST} based on the user device type to the ST users, whereas the PR users are always guaranteed with QoE level higher than QoE_{ST} . The deviation in the mean delivered QoE is due to the following two reasons: 1) in the proposed approach, the resource allocation is based on the $O.22$ score of the ITU-T P.1203 standard, whereas the evaluation of the obtained final QoE prediction score is based on the ITU-T P.1203 $O.46$, which considers more QoE influencing factors than $O.22$; 2) large quantization levels of video representation bitrates does not allow for reaching exactly a desired level of quality. For example, in case of the FHD users, significant shifts in delivered QoE can be observed while varying QoE_{ST} , which is due to the significant difference between the bitrates (quantization level) of the representations. From Table VI it arises that the zero-rated data rate approach can neither guaranteed delivered QoE according to device type nor according to the CoS. Indeed, higher delivered QoE can only be observed in the case of users with SD device types (not according to the CoS), as the control is performed only on the assigned throughput. Specifically, users having FHD devices receive lower QoE. For example, in the case of PR FHD users, zero-rated QoE delivers mean QoE as high as 4.2, whereas zero-rated data rate delivers mean QoE maximum up to 3.6 MOS. In general, the consequence is that the higher the video resolution, the lower the quality.

Table VII shows the resulting assigned throughput for the proposed algorithm for the different CoS and device types. For the $QoE_{ST} \leq 3.6$, the zero-rated QoE approach utilizes less total throughput and delivers better QoE as compared to the zero-rated data rate approach. Whereas for $QoE_{ST} > 3.6$, the zero-rated QoE approach utilizes total throughput comparable to the zero-rated data rate approach. For instance, when $QoE_{ST} = 3.6$, the total throughput utilized by the zero-rated approach is 65.60 Mbps, which is comparable to zero-rated data rate approach $T_{ZR} = 1100$ Kbps (total throughput utilized is 66 Mbps). However, it can be seen from the Table VI that the zero-rated QoE delivers higher mean QoE for $QoE_{ST} = 3.6$ as compared to zero-rated data rate approach at $T_{ZR} = 1100$ Kbps. This is due to the fact that zero-rated QoE approach allocates the resources

TABLE VII
THROUGHPUT DISTRIBUTION IN ZERO-RATED QOE APPROACH VS. ZERO-RATED DATA RATE APPROACH

Zero-rated QoE approach								**Zero-rated Data rate approach**	
QoE_{ST}	Mean Throughput (Mbps)						Total Throughput (Mbps)	T_{ZR}	Total Throughput (Mbps)
	PR FHD	PR HD	PR SD	ST FHD	ST HD	ST SD			
2	0.51	0.37	0.18	0.40	0.19	0.18	16.34	500	30
2.2	0.55	0.37	0.18	0.53	0.20	0.18	17.89	600	36
2.4	0.55	0.67	0.18	0.53	0.35	0.18	23.25	700	42
2.6	0.89	0.67	0.26	0.62	0.35	0.18	26.57	800	48
2.8	0.89	0.67	0.35	0.62	0.44	0.18	29.16	900	54
3	1.16	0.80	0.44	0.88	0.63	0.18	38.06	1000	60
3.2	1.31	1.06	0.59	0.99	0.63	0.35	45.32	**1100**	**66**
3.4	2.02	1.25	0.65	0.99	0.68	0.35	52.04	1200	72
3.6	**2.41**	**1.84**	**0.69**	**1.46**	**0.74**	**0.63**	**65.60**	1300	78
3.8	3.31	1.84	0.69	2.02	1.17	0.63	84.51	1400	84
4	3.31	1.84	0.69	2.23	1.33	0.75	90.19	1500	90

considering the device type and CoS, whereas the zero-rated data rate approach provides equal throughput to all users, i.e., to have same QoE level the users with FHD resolution devices require higher amount of resources at network level as compared to user with SD devices. Moreover, allocation of higher throughput than required (higher video bitrates and higher resolution, e.g., HD) to SD users may not improve the user perceived quality further, which is validated in [50]. Thus, in the zero-rated QoE approach, the network resources are saved while allocating network resources to users with SD devices according to the device requirements, whereas more network resources are given to FHD users according to the device requirements. Similarly, the zero-rated QoE approach also considers CoS, where higher throughput is allocated to the PR users as compared to ST users to deliver better mean QoE.

Fig. 3 provides a comparison of the two approaches in terms of the QoE delivered when using a similar amount of the total network resources. In this figure, we are also providing the 95% confidence interval. For the zero-rated QoE approach, QoE_{ST} = 3.6 is select, which utilizes a total throughput equal to 65.659 Mbps. This is comparable to the zero-rated data rate approach at T_{ZR} = 1100 Mbps, which utilizes a total throughput of 66 Mbps. The zero-rate QoE approach delivers an overall QoE level of 4.1 to PR users as compared to 3.6 for the zero-rated data rate approach. Whereas for the ST users, the overall QoE delivered by the zero-rated QoE approach is 3.8, which is slightly higher than the QoE delivered to ST users by the zero-rated data rate approach that is of 3.6. However, the QoE level delivered to ST users above 3 MOS is still acceptable since the users who pay less expect less [38]. Moreover, the zero-rated QoE approach delivers QoE higher than 4 MOS to PR users of all device types while zero-rated data rate approach delivers lower QoE to users with high-resolution devices (for FHD 3.2 and 3.7 for HD) while higher QoE is delivered to the users with low-resolution device type (SD) up to 3.9. Moreover, the zero-rated data rate approach delivers the same QoE to PR and ST users, regardless of CoS.

Fig. 4 provides a comparison of the two approached in terms of the fairness index using Eq. (7). It shows that the zero-rated QoE approach delivers better fairness within a class as compared to the alternative approach. Thus, the

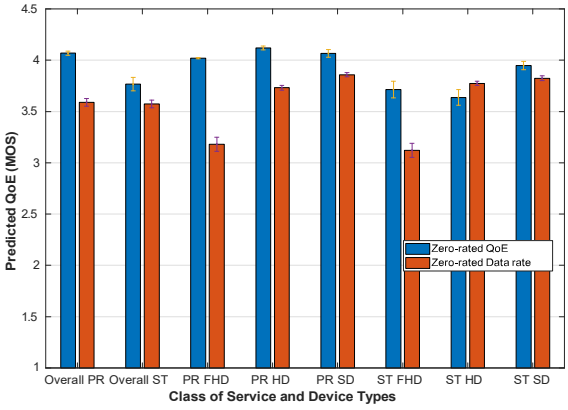

Fig. 3. Comparison in terms of delivered QoE at QoE_{ST} = 3.6 MOS and T_{ZR} = 1100 Kbps. This graphs shows also the 95% confidence interval.

users belonging to the same CoS have more or less the same QoE level (independently from the device type) in the zero-rated QoE approach, which leads to a fairer distribution of the network resources within the class. Whereas in the zero-rated data rate approach, the users belonging to the same CoS receives different QoE (depending on the device type), i.e., on the same throughput for instance at T_{ZR} = 1100 Kbps, users with SD device receive higher QoE as compared to users with higher resolution devices. The results validate the findings of the study in [46], according to which equal distribution of throughput (Fairness based on the Jane Fairness index) doesn't ensure fairness in terms of QoE. Moreover, in the zero-rated data rate approach, the fairness improves as the T_{ZR} increases, which is due to the improvement in the delivered QoE to the high-resolution devices (FHD and HD). The minimum and maximum fairness delivered by the zero-rated QoE approach to the PR users are 0.8059 and 0.99, respectively. While for ST users, the zero-rated QoE has minimum and maximum fairness up to 0.7163 and 0.9846 respectively. On the other hand, the minimum and maximum fairness delivered by the zero-rated data rate approach to PR users are 0.6613 and 0.9281, respectively, while minimum and maximum fairness for the ST users are 0.6608 and 0.8965 respectively. For a fairer comparison, we focus on the results corresponding to

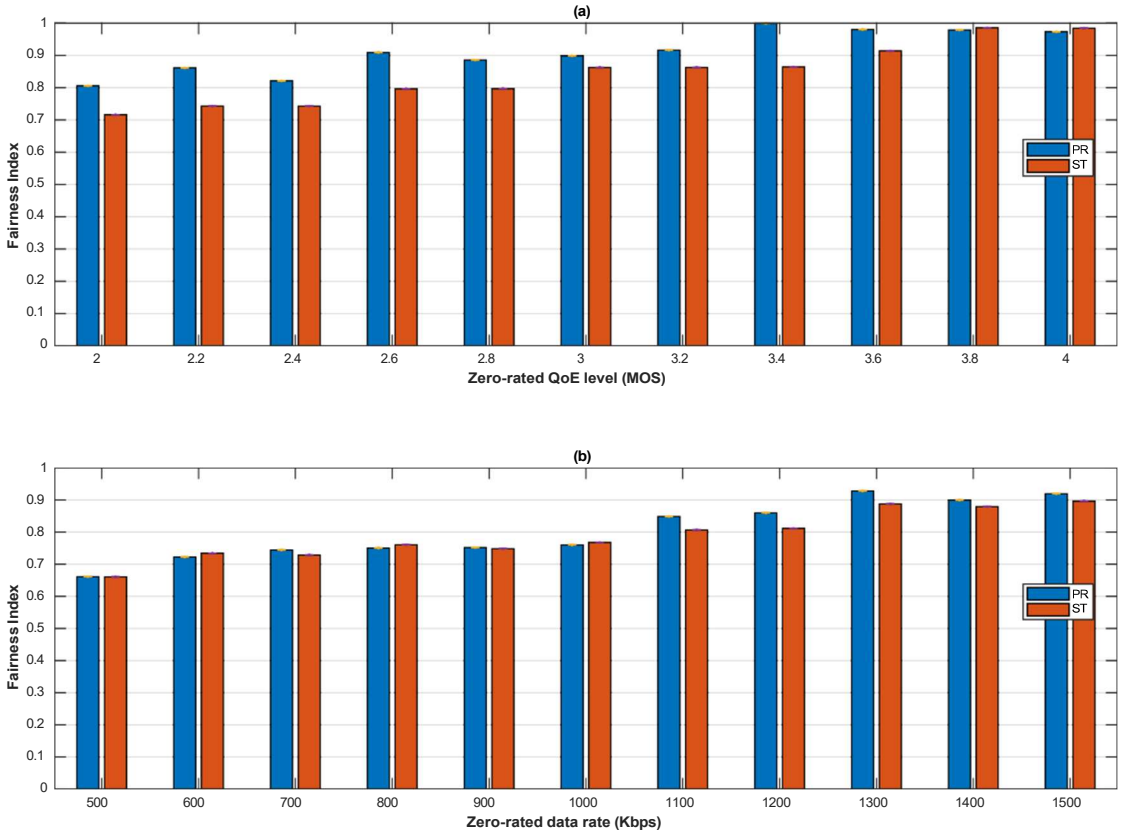

Fig. 4. Comparison of the two algorithms in terms of fairness by varying the target QoE and throughput: (a) Zero-rate QoE approach and (b) Zero-rated data rate approach.

QoE_{ST} = 3.6 (for the proposed approach) and T_{ZR} = 1100 Kbps (for the alternative one), as the resources utilized with these settings are comparable. It can be seen that the zero-rated QoE approach delivers better fairness to both PR and ST users, which are equal to 0.9804 and 0.9142, respectively, whereas with the alternative approach the fairness values are of 0.8486 and 0.8068 for the PR and ST users, respectively. Thus, the proposed approach is significantly fairer than the alternative solution.

VI. Conclusion

In this work, we propose a reference architecture for the collaborative QoE management of the OTT video streaming in 5G networks with the inclusion of the latest standardization activities at 3GPP and MPEG DASH. We also defined the information to be exchanged and interfaces among QoE-aware network elements and client-side OTT application for the QoE monitoring, network-assisted video quality adaptation and QoE-aware network resource management. Then we propose a zero-rated QoE approach with the context-aware QoE based network resource allocation algorithm. In the resource allocation algorithm, the user's CoS, device resolution, and video characteristics are also considered. The conducted simulation results provide the comparative analysis of the proposed zero-rated QoE approach with the zero-rated data rate approach on multiple scales: QoE delivered, QoE fairness among the users, delivered video quality, the evolution of the generated profit over the time and effective network resource utilization. The proposed approach has proven to outperform the zero-rated data rate approach on all scales leading to effective QoE-aware network resource management.

The future work will also be centered around the evaluation of the approach in terms of scalability. Indeed, the proposed algorithm requires performing continuous monitoring of the QoE for each user, which may be quite complex in terms of generated control traffic and computation overhead for quality prediction. A possible solution, still to be verified, is to adopt more simplified models that make use of only metadata for the considered services, without computing the quality of each video representation as performed with the current version of the proposed zero-rated QoE approach.

Acknowledgment

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

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2016-2021 - White paper," 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[2] L. Skorin-Kapov, M. Varela, T. Hoßfeld, and K.-T. Chen, "A survey of emerging concepts and challenges for qoe management of multimedia services," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 2s, p. 29, 2018.

[3] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.

[4] P. Le Callet, S. Mo¨ller, A. Perkis *et al.*, "Qualinet White Paper on Definitions of Quality of Experience (2012)," in *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, Lausanne, Switzerland, Version 1.2, March 2013.

[5] S. Barakovic´ and L. Skorin-Kapov, "Survey and Challenges of QoE Management Issues in Wireless Networks," *Journal of Computer Networks and Communications*, vol. 2013, 2013.

[6] J. Seppa¨nen, M. Varela, and A. Sgora, "An autonomous QoE-driven network management framework," *Journal of Visual Communication and Image Representation*, vol. 25, no. 3, pp. 565–577, 2014.

[7] A. Floris, A. Ahmad, and L. Atzori, "QoE-aware OTT-ISP collaboration in service management: Architecture and approaches," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 2s, p. 36, 2018.

[8] S. Petrangeli, J. V. D. Hooft, T. Wauters, and F. D. Turck, "Quality of experience-centric management of adaptive video streaming services: Status and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 2s, p. 31, 2018.

[9] I. 23009-1:2014, "Information Technology–Dynamic Adaptive Streaming Over HTTP (DASH)–Part 1: Media Presentation Description and Segment Formats," 2014.

[10] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of http adaptive streaming," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[11] D. Z. Rodr´ıguez, Z. Wang, R. L. Rosa, and G. Bressan, "The impact of video-quality-level switching on user quality of experience in dynamic adaptive streaming over HTTP," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, 2014.

[12] I. 23009-5:2017, "Dynamic adaptive streaming over HTTP (DASH)–Part 5: Server and network assisted DASH (SAND)," 2017.

[13] A. Bentaleb, A. C. Begen, R. Zimmermann, and S. Harous, "Sdnhas: An sdn-enabled architecture to optimize qoe in http adaptive streaming," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2136–2151, 2017.

[14] A. M. Kakhki, F. Li, D. Choffnes, E. Katz-Bassett, and A. Mislove, "Bingeon under the microscope: Understanding t-mobiles zero-rating implementation," in *Proceedings of the 2016 workshop on QoE-based Analysis and Management of Data Communication Networks*. ACM, 2016, pp. 43–48.

[15] Y. Wang, P. Li, L. Jiao, Z. Su, N. Cheng, X. S. Shen, and P. Zhang, "A data-driven architecture for personalized qoe management in 5g wireless networks," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 102–110, 2017.

[16] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5g network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, 2014.

[17] M. Gramaglia, I. Digon, V. Friderikos, D. von Hugo, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Flexible connectivity and qoe/qos management for 5g networks: The 5g norma view," in *Communications Workshops (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 373–379.

[18] M. Jiang, M. Condoluci, and T. Mahmoodi, "Network slicing management & prioritization in 5g mobile systems," in *European wireless*, 2016, pp. 1–6.

[19] S. Dutta, T. Taleb, and A. Ksentini, "Qoe-aware elasticity support in cloud-native 5g systems," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[20] C. Ge, N. Wang, S. Skillman, G. Foster, and Y. Cao, "Qoe-driven dash video caching and adaptation at 5g mobile edge," in *Proceedings of the 3rd ACM Conference on Information-Centric Networking*. ACM, 2016, pp. 237–242.

[21] G. Cofano, L. D. Cicco, T. Zinner, A. Nguyen-Ngoc, P. Tran-Gia, and S. Mascolo, "Design and Performance Evaluation of Network-assisted Control Strategies for HTTP Adaptive Streaming," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 3s, pp. 42:1–42:24, Jun. 2017.

[22] E. Khorov, A. Krasilov, M. Liubogoshchev, and S. Tang, "Sebra: Sand-enabled bitrate and resource allocation algorithm for network-assisted video streaming," in *2017 IEEE 13th Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2017, pp. 1–8.

[23] M. Varela, P. Zwickl, P. Reichl, M. Xie, and H. Schulzrinne, "From Service Level Agreements (SLA) to Experience Level Agreements (ELA): The challenges of selling QoE to the user," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1741–1746.

[24] A. Ahmad, A. Floris, and L. Atzori, "Ott-isp joint service management: a customer lifetime value based approach," in *Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on*. IEEE, 2017, pp. 1017–1022.

[25] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5g wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[26] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[27] T. HoBfeld, M. Varela, P. E. Heegaard, and L. Skorin-Kapov, "Observations on emerging aspects in qoe modeling and their impact on qoe management," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–6.

[28] L. Atzori and M. Lobina, "Speech playout buffering based on a simplified version of the itu-t e-model," *IEEE Signal Processing Letters*, no. 11, pp. 382–385, 2004.

[29] P. Hosein, W. Choi, W. Seok *et al.*, "Disruptive network applications and their impact on network neutrality," in *Advanced Communication Technology (ICACT), 2015 17th International Conference on*. IEEE, 2015, pp. 663–668.

[30] H. H. Gharakheili, "Perspectives on net neutrality and internet fast-lanes," in *The Role of SDN in Broadband Networks*. Springer, 2017, pp. 5–22.

[31] A. Ahmad, A. Floris, and L. Atzori, "Towards qoe monitoring at user terminal: A monitoring approach based on quality degradation," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 2017, pp. 1–6.

[32] A. Ahmad, L. Atzori, and M. G. Martini, "Qualia: A multilayer solution for QoE passive monitoring at the user terminal," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[33] A. Ahmad, A. Floris, and L. Atzori, "Timber: An sdn based emulation platform for qoe management experimental research," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–6.

[34] 3GPP, "3GPP TS 23.501 V15.1.0 (2018-03); Technical Specification Group Services and System Aspects; System Architecture for the 5G System ; Stage 2; (Release 15)," 2018.

[35] A. Ahmad, A. Floris, and L. Atzori, "Towards information-centric collaborative QoE management using SDN," in *2019 IEEE Wireless Communications and Networking Conference (WCNC) (IEEE WCNC 2019)*, Marrakech, Morocco, Apr. 2019.

[36] P. Paudyal, F. Battisti, and M. Carli, "Impact of video content and transmission impairments on quality of experience," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 16 461–16 485, 2016.

[37] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing effect sizes of influence factors towards a qoe model for http adaptive streaming," in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*. IEEE, 2014, pp. 111–116.

[38] A. Ahmad, A. Floris, and L. Atzori, "QoE-centric service delivery: A collaborative approach among OTTs and ISPs," *Computer Networks*, vol. 110, pp. 168–179, 2016.

[39] I. Abdeljaouad and A. Karmouch, "Monitoring iptv quality of experience in overlay networks using utility functions," *Journal of Network and Computer Applications*, vol. 54, pp. 1–10, 2015.

[40] Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. C. Begen, and D. Oran, "Probe and adapt: Rate adaptation for http video streaming at scale," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 719–733, 2014.

[41] E. Liotou, K. Samdanis, E. Pateromichelakis, N. Passas, and L. Merakos, "Qoe-sdn app: A rate-guided qoe-aware sdn-app for http adaptive video streaming," *IEEE Journal on Selected Areas in Communications*, 2018.

[42] 3GPP, "3GPP TS 38.211 V15.5.0 (2019-03); Technical Specification Group Radio Access Network; NR; Physical channels and modulation; (Release 16)," 2019.

[43] International Telecommunication Union, “Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport,” 2017.

[44] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Gring, and B. Feiten, “A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1,” in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. Erfurt: IEEE, May 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7965631/

[45] W. Robitza, S. Gring, A. Raake, D. Lindegren, G. Heikkil, J. Gustafsson, P. List, B. Feiten, U. Wstenhagen, M.-N. Garcia, K. Yamagishi, and S. Broom, “HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 Open Databases and Software,” in *9th ACM Multimedia Systems Conference*, Amsterdam, 2018.

[46] T. Hoßfeld, L. Skorin-Kapov, P. E. Heegaard, and M. Varela, “Definition of QoE Fairness in Shared Systems,” *IEEE Communications Letters*, vol. 21, no. 1, pp. 184–187, 2017.

[47] S. Lederer, C. Mu¨ller, and C. Timmerer, “Dynamic adaptive streaming over http dataset,” in *Proceedings of the 3rd Multimedia Systems Conference*. ACM, 2012, pp. 89–94.

[48] L. De Cicco, V. Caldaralo, V. Palmisano, and S. Mascolo, “Tapas: a tool for rapid prototyping of adaptive streaming algorithms,” in *Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming*. ACM, 2014, pp. 1–6.

[49] 3GPP, “3GPP TS 36.101 V16.1.0 (2019-04); Technical Specification Group Services and System Aspects; User Equipment (UE) radio transmission and reception ; Stage 2; (Release 16),” 2019.

[50] N. Abis, A. Floris, S. Argyropoulos, L. Atzori, and A. Raake, “I have to switch the terminal: evaluating the impact on video quality perception,” in *Communications (ICC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 6977–6982.