

Research Article

A Framework for QoE-Aware 3D Video Streaming Optimisation over Wireless Networks

Ilias Politis, Asimakis Lykourgiotis, and Tasos Dagiuklas

CONES Research Group, Hellenic Open University, 26335 Patras, Greece

Correspondence should be addressed to Ilias Politis; ilpolitis@eap.gr

Received 6 October 2015; Revised 5 February 2016; Accepted 29 February 2016

Academic Editor: Ansgar Gerlicher

Copyright © 2016 Ilias Politis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The delivery of three-dimensional immersive media to individual users remains a highly challenging problem due to the large amount of data involved, diverse network characteristics, and user terminal requirements, as well as user's context. This paper proposes a framework for quality of experience-aware delivering of three-dimensional video across heterogeneous wireless networks. The proposed architecture combines a Media-Aware Proxy (application layer filter), an enhanced version of IEEE 802.21 protocol for monitoring key performance parameters from different entities and multiple layers, and a QoE controller with a machine learning-based decision engine, capable of modelling the perceived video quality. The proposed architecture is fully integrated with the Long Term Evolution Enhanced Packet Core networks. The paper investigates machine learning-based techniques for producing an objective QoE model based on parameters from the physical, the data link, and the network layers. Extensive test-bed experiments and statistical analysis indicate that the proposed framework is capable of modelling accurately the impact of network impairments to the perceptual quality of three-dimensional video user.

1. Introduction

Media-friendly Future Internet needs to cope with an increased demand for streaming applications and three-dimensional (3D) media by modifying and optimising existing protocols for streaming over a hybrid network environment. Accessing the Internet easily, any time, anywhere, and with any device, has changed user's Internet consumption amount and behaviour, especially on media consumption [1]. Traditionally web surfing and file sharing were the dominant applications, whereas today real-time streaming and social media applications (e.g., YouTube, Facebook, etc.) constitute the major portion of the Internet traffic. The change in Internet usage patterns and the increasing needs of novel applications (high performance, availability, security, etc.), together with end users' increasing quality of experience (QoE) expectations, open new challenges for the Future Internet due to increasing popularity of real-time communication, online social networks, and heterogeneity of user devices [2–4].

It is evident that due to recent advances in video acquisition techniques, coding, delivery, and displaying, 3D video has sprouted in the consumer domain through a range of

applications and services which provide enhanced visual experience for the viewers. Recently, the research on 3D video coding and transportation intensified, flared by enhancements and updates on coding strategies, such as H.264/SVC (Scalable Video Coding) [5], H.264/MVC (Multiview Video Coding) [6], and High Efficiency Video Coding (HEVC) [7], as well as the wide deployment of LTE (Long Term Evolution) systems for high-speed data delivery to both wireless and mobile users [8].

Based on the different representation formats for stereoscopic and multiview videos, different coding schemes have evolved over time to compress such media efficiently. Pure colour based (e.g., multiview, stereo L-R) or depth based formats (e.g., colour-plus-depth or Layered Video plus Depth) are usually compressed using the legacy block based coding standards, such as MPEG-4 Part 10/H.264 Advance Video Coding (AVC), or its extensions, such as SVC [9] or MVC [10]. AVC has also introduced new profiles to support coding of stereoscopic videos, such as Stereoscopic High Profile [11]. Both AVC derived codecs, namely, SVC and MVC, get their base layer fully compliant with AVC specifications in order to be compliant with legacy deployments. The purpose of SVC is

to provide a universal media bit-stream that can be decoded by multiple decoders of different capacities to produce the reconstructed media at different states. SVC also provides dynamic adaptation to a diversity of networks, terminals, and formats. This is extremely useful for simple adaptation of transmission and efficient storage and is beneficial for transmission services with uncertain resolution, channel conditions, and device types. The MVC standard, on the other hand, in order to improve the overall rate-distortion performance, exploits the inter-view redundancies through the Disparity Compensated Prediction (DCP) [12]. Nonetheless, the multiview bit rate increases to unmanageable levels as the number of source cameras increases. Therefore, it is impractical to encode and transport the entire set of views required to be displayed on most multiview displays, particularly when dealing with heterogeneous wireless access networks. Yet, it remains unclear how all these technologies affect the viewer's perception of 3D video [13, 14]. Apparently, it is still unrealistic to try and estimate the QoE of such application, without resorting to full subjective evaluations.

Any 3D video coding and delivery system will ultimately need to be measured in terms of user's satisfaction and perceived experience. QoE can be defined as the overall acceptability of an application or service strictly from the users point of view. It is a subjective measure of end-to-end service performance from the user perspective and it is an indication of how well the network meets the users needs [15]. Encompassing many different aspects, QoE is riveted on the true feelings of end users when they watch streaming video, listen to digitized music, and browse the Internet through a plethora of methods and devices [16]. There is a real challenge in creating models which will accurately perform learning to model service behaviours by taking into account parameters such as arrival pattern request, service time distributions, I/O system behaviours, and network usage [17]. The aim of such attempts is to estimate QoE for both resource-centric (utilization, availability, reliability, and incentive) and user-centric (response time and fairness) environments over certain predefined Quality of Service (QoS) and QoE thresholds. Evidently, there is still no concrete evidence to prove the correlation between QoS and QoE, particularly for 3D video. Nevertheless, there are several attempts to define the relationship of human perception of video quality to the inherently objective network conditions, and the majority of this work concludes that such a relationship cannot be linear [18, 19]. Specifically, authors in [20] proved the importance of the impact that network delivery speed and latency have on the human satisfaction but also determined that, in a heterogeneous networking environment, neither bandwidth nor latency is sufficient as attributes that indicate the range of issues, which render a service as nonappealing to the end user.

This paper describes a framework for QoE-aware delivering of 3D video across heterogeneous wireless networks. Towards this end, a synergy is proposed between the LTE EPC architecture and IEEE 802.21 protocol that will ensure real-time control and monitoring of key performance indicators and parameters relative to the perceived 3D QoE across different layers and entities. The proposed framework includes

a QoE controller module housing a machine learning- (ML-) based decision making engine, which trains data collected from the user and the network planes in order to model, as accurate as possible, the perceptual 3D video quality. It needs to be underlined that the proposed architecture does not require any alteration of the standardised LTE EPC interfaces; hence, it could be smoothly integrated with the current mobile operators' deployments. The paper investigates machine learning-based techniques for producing an objective QoE model of network related impairments. The proposed QoE model is a function of parameters collected not only from the application layer, but also from the underlying layers. Previous efforts on modelling QoE for 3D video have been based on measuring the end-to-end packet loss and determining the perceived QoE, a process that would require feedback from the receiving end, which in the case of wireless networks may never arrive or may arrive too late for the decision engine. This research investigates a new methodology for monitoring network imposed impairments to the perceived video quality that will result in more accurate QoE models. Specifically, a set of QoS parameters is collected from different points in the delivery chain (i.e., bit error rate from physical layer, MAC layer load, and network layer delay/jitter), which account for the overall number of lost packets. Three classification schemes with different characteristics have been studied in order to identify the most appropriate method for modelling 3D video quality due to network impairments.

The rest of the paper is organised as follows. Section 2 briefly presents the concept of LTE EPC and its main entities that are relevant to the purposes of this study. Section 3 provides a detailed description of the proposed media-aware framework and focuses on the synergy between the Media-Aware Proxy, the IEEE 802.21 protocol, and the QoE controller, which enables collecting QoE related key performance indicators across multiple layers and network components, the training of these data, and the modelling of the perceived 3D video QoE. Additionally, in this section the three ML-based classification models are introduced and their main differences are explained. The setup of the experimental test-bed is described in Section 4. Moreover, the section defines the physical, MAC, and network layer parameters, whose impact on the perceptual quality of 3D video is under study. The collected Mean Opinion Score (MOS) from the subjective experiments, along with a detailed statistical analysis, is presented in Section 5. Based on the measured 3D MOS, Section 6 compares the performance of the three machine learning methods in terms of precision and accuracy, in order to determine the most suitable classification scheme for implementation as part of the QoE controller. Finally, Section 7 concludes the paper and highlights the future aims of this research.

2. 3GPP EPS Architecture

The technical realization of next generation mobile networks requires that complementary wireless technologies are integrated in order to provide ubiquitous multimedia services "anywhere, any time, and on any device." In the

recent years, 3GPP's Evolved Packet System (EPS) [21] is increasingly deployed by mobile operators in order to satisfy the requirement for uninterrupted, high quality, real-time multimedia services across wireless access networks. EPS consists of the LTE wireless access, which is forming the Evolved UTRAN (E-UTRAN) as the lower part of EPS and an IP connectivity control platform called Evolved Packet Core (EPC) as the upper part of EPS, which enable wireless access networks diversity (i.e., LTE, UMTS, WiMax, WiFi, etc.). EPS is relying on standardized routing and transport mechanisms of the underlying IP network. In parallel, EPS provides capabilities for coexistence with legacy systems and migration to the EPS, while it supports functionalities for access selection based on operators policies, user preferences, and networking conditions. One of the most important features of EPS is its ability to support simultaneous active Packet Data Network (PDN) connections for the same user equipment, whilst maintaining the negotiated QoS across the whole system.

3GPP has proposed the EPC in an effort to support higher data rates, lower latency, packet optimisation, and multiple radio access technologies [22]. As part of EPS, EPC is an all-IP architecture that fulfils those requirements and supports the characteristics of E-UTRAN. The advantages of EPC, as opposed to the legacy GPRS architecture, are a more clear depiction of control and data planes, a more simplified architecture with a single core network, and, finally, the full assumption of the IETF protocols. Therefore, EPC allows for a truly converged packet core for trusted 3GPP networks (i.e., GPRS, UMTS, LTE, etc.), non-3GPP networks (i.e., WiMAX), and untrusted networks (i.e., WLAN). Additionally, EPC maintains seamless mobility and consistent and optimised services provisioning independent of the underlying access network.

The proposed framework for QoE-aware delivery of 3D video content over heterogeneous wireless networks relies on 3GPP EPC. Therefore, a description of the EPC modules related to the study follows. The functionalities of the discussed modules are enhanced as part of the proposed framework in order to support seamless connectivity, transparent real-time adaptation of 3D video streaming, and QoE-aware monitoring and management.

2.1. Key EPC Modules. The relevance to the proposed framework EPC modules [22] is depicted in Figure 1. Specifically, these architectural modules and functions include the following:

- (i) Home Subscriber Server (HSS) is the main database of the EPC responsible for storing subscriber's information that may include profile description, several identification information, and information regarding the subscriber's established sessions. The module interfaces with the Mobility Management Entity and the Authentication, Authorization, and Accounting server (implemented as a Diameter server [23]).
- (ii) Mobility Management Entity (MME) is the module that performs the mobility management and bearer management and establishment. It is also responsible

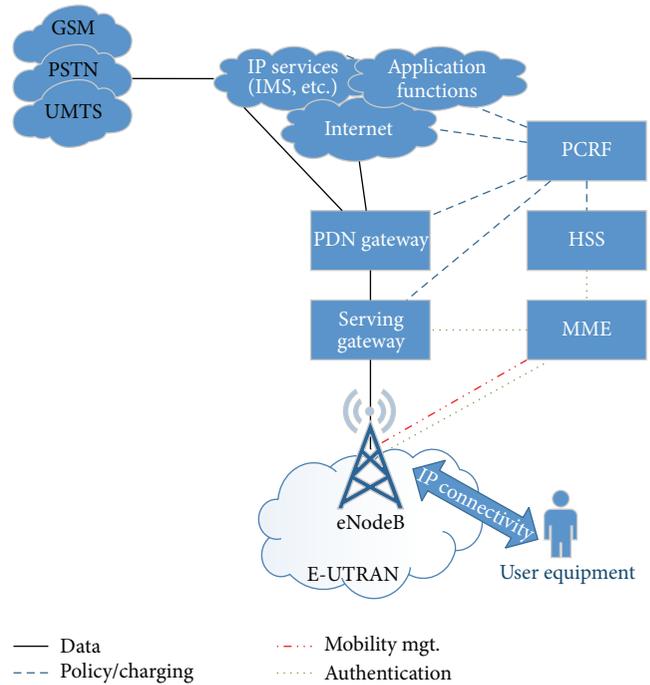


FIGURE 1: Relevant EPC architectural elements.

for performing mobility functions during handovers between 2G or 3G 3GPP access networks. The module has interfaces to the Serving-GW, the eNodeB, and the HSS.

- (iii) Serving Gateway (Serving-GW) acts as the gateway where the interface towards E-UTRAN terminates. Each piece of user equipment (UE) associated with the EPS is served by a single Serving-GW. In terms of implementation, both the Serving-GW and the PDN-GW may be dwelling in a single physical component.
- (iv) Packet Data Network Gateway (PDN-GW) is the gateway which terminates the interface towards the packet data networks (e.g., IMS [24], Internet). In case where the UE is connected with different IP services, several PDN-GW are associated with this UE. It performs UE IP allocation, policy enforcement, packet filtering per user, and packet screening. Moreover, PDN-GW acts as the mobility anchor between 3GPP and non-3GPP access. It has interfaces to the Serving-GW, the Evolved Packet Data Gateway (ePDG), which is responsible for interworking between the EPC and untrusted non-3GPP networks, such as WiFi, interface to the 3GPP AAA server, the trusted non-3GPP access gateway, and the Policy and Charging Rules Function.
- (v) Policy and Charging Rules Function (PCRF) performs QoS control, access control, and charging control. The function is policy based and authorizes bearer and session establishment and management. A single PCRF is assigned to all the connections of a subscriber. It interfaces with Policy and

Charging Enforcement Function (PCEF) within the PDN-GW and the Bearer Binding and Event Reporting Functions (BBERFs) within the trusted/untrusted 3GPP/non-3GPP access gateways and the HSS.

- (vi) Evolved-UTRAN (E-UTRAN), as explained above, is a simplified radio access network architecture comprised of evolved NodeBs (eNodeBs). In terms of functionality E-UTRAN supports radio resource management (RRM) and selection of MME and is responsible for routing user data to the Serving-GW and for performing scheduling, measurements, and reporting.

The described EPC modules have been implemented as part of the *OpenEPC* testing platform [25], and they have been extended within the context of this work in order to support QoE-aware multimedia delivery, as described in the rest of the paper.

3. Proposed Media-Aware Architecture

The media-aware architecture proposed in this paper aims to provide seamless end-to-end services with ubiquitous use of the heterogeneous technologies. In the centre of this architecture lies a QoE controller, which facilitates a ML-based model of the quality of the perceived video experience. The learning process is based on a collection of key performance indicators from across the network architecture and from different layers. This information is stored in a central database, as described in the rest of this section. Evidently, the proposed framework is required to provide seamless, uninterrupted, and QoE-aware video services across heterogeneous link layer interfaces and user devices. Therefore, the proposed media-aware framework integrates a novel entity named Media-Aware Proxy (MAP) with the IEEE 802.21 Media Independent Handover (MIH) protocol [26]. In order to support seamless handover and IP layer mobility, the proposed framework supports also the Proxy Mobile IP; however, this is beyond the scope of the current research.

This framework is tightly coupled with the LTE EPC functions and modules [22]; however, it neither interrupts the inherent process of EPC nor violates its protocols, as it lies just before the core network. The synergy of MAP and MIH along with a QoE controller, which acts as the QoE-aware decision making function, allows for real-time video stream adaptation (i.e., intelligent quality layer dropping or stream prioritization) based on multiple physical, network, and application layer parameters, collected from both the mobile terminal and the access network side. An overview of the proposed architecture is illustrated in Figure 2.

3.1. Media-Aware Proxy. MAP is defined as a transparent user-space module responsible for low-delay adaptation and filtering of scalable video streams [27]. In conjunction with the QoE controller, which models QoE and informs MAP about the predicted level of the perceived quality, the latter is able to either drop or forward packets that carry specific layers of a stream to the receiving video users. In detail, MAP is a network function based on the Media-Aware Network

Element (MANE) standard [28, 29]. Briefly, MANE can be considered as either a middle box or an application layer getaway capable of aggregating or thinning RTP streams by selectively dropping packets that have the less significant impact on the user's video experience. As such, MANE has been proposed as an intermediate system that is capable of receiving and depacketising RTP traffic in order to customise the encapsulated network abstraction layer units, according to client's and access network's requirements. Within the context of the proposed media-aware architecture, MAP's role is twofold. Firstly, it acts as a central point of decision in order to overcome networking limitations imposed by firewalls and Network Address Translation (NAT) protocol that are extensively used in real life networks. Secondly, it receives and parses RTP streams and customises the streaming according to the video client's requirements and network conditions, based on the QoE-aware decision engine of the QoE controller.

In particular, MAP, which is designed to run in Linux kernel level in order to ensure minimum impact on the end-to-end delay, acts as a transparent proxy of the mobile client and it is responsible for parsing the packets that are destined for all mobile users over multiple ports. Each received packet is forwarded in a queue and its header is parsed by an RTP parser process in order to identify the embedded video related information, without changing the header's fields. Moreover, MAP has the responsibility to store this information in media independent information service (MIIS) database. MAP is designed as a MIH-aware entity; hence, it can directly store information required from the QoE model to the MIIS database. Specifically, MAP regularly updates the database with all client IDs (i.e., video service, video view ID, number of layers, incoming data rate, etc.); thus, it enables the QoE controller to train the collected data and predict the perceptual video quality of experience. The timely output of the ML-based QoE model will be utilised by MAP in order to maximise the perceived video QoE by selecting the appropriate stream optimisation method. The abovementioned algorithmic process is shown in Figure 3.

3.2. IEEE 802.21 Overview. While the main purpose of IEEE 802.21 standard is to enhance the handover performance in heterogeneous networks [26], it also supports the important aspect of link adaptation. A resource-intensive multimedia application like video streaming requires link adaptation in order for the network provider to offer the maximum level of video experience. As a result IEEE 802.21 standard can be employed to support cooperative use of information available at both the mobile node and the network infrastructure. This framework introduces a new entity called MIH Function (MIHF), which resides within the protocol stack of the network elements and particularly between the link layer and the upper layers. Any protocol that uses the services provided by the MIHF is called a MIH user.

As shown in Figure 4, MIHF provides three types of services, the media independent event service (MIES), the media independent command service (MICS), and the media independent information service (MIIS). The MIES monitors link layer properties and initiates related events in order to

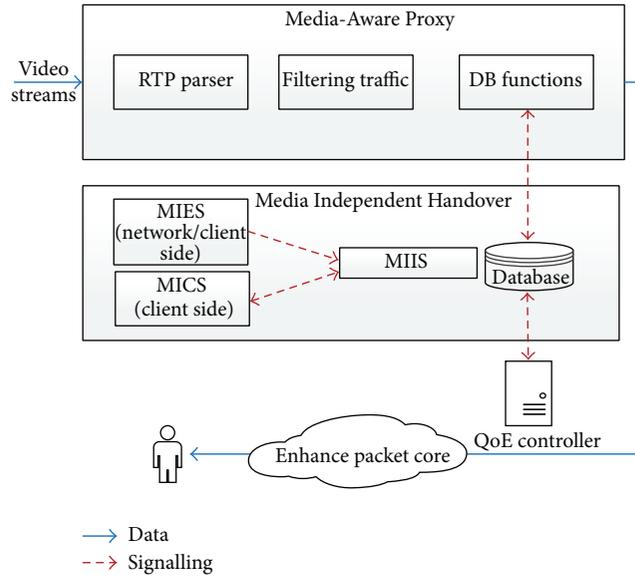


FIGURE 2: Proposed media-aware framework overview.

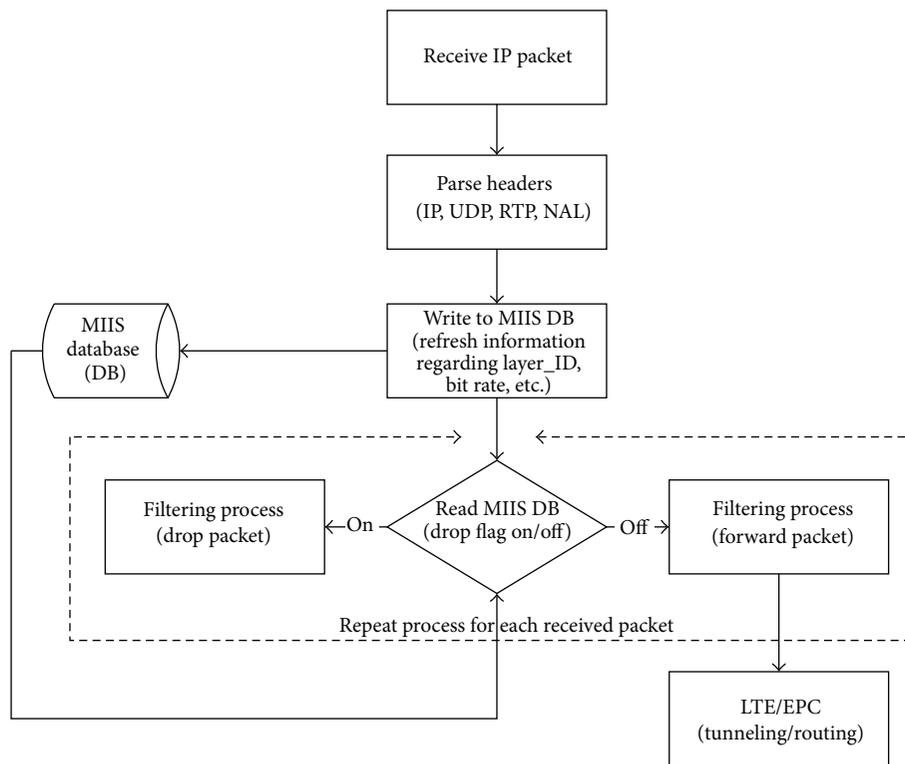


FIGURE 3: Algorithmic representation of Media-Aware Proxy.

inform both local and remote users. The MICS provides a set of commands for the MIH users to control link properties and interfaces. Finally, the MIIS provides information of static nature for the candidate access networks such as frequency bands, maximum data rate, and link layer address.

On the other hand, there is the 3GGP approach which defines a new component called access network discovery

and selection function (ANDSF) [30], which in conjunction with the event reporting function (ERF) aims to facilitate information exchange for the network selection process. However, this approach is not as exhaustive and comprehensive as the MIIS and can not support in depth monitoring of the link layer condition. For this reason, in the current work, the integration of MIH in the EPC architecture to support

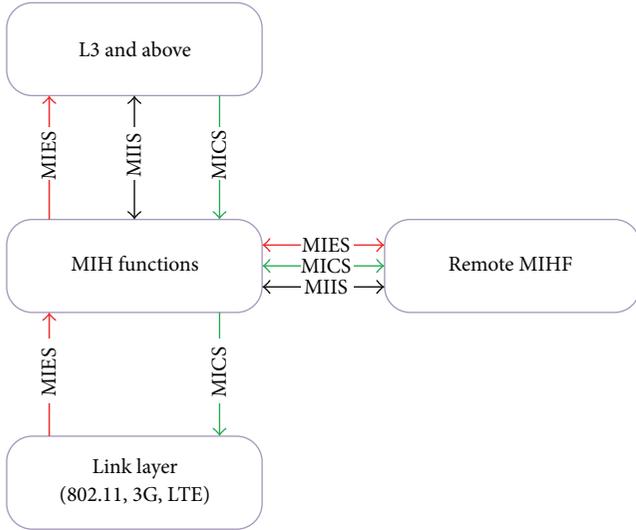


FIGURE 4: Illustration of overall MIH architecture and services.

link layer monitoring is proposed and the related signalling is presented. Furthermore, the available set of monitored characteristics is extended in order to include the application layer and, hence, to encompass video related attributes.

In order for the MIH framework to be effectively deployed in the EPC architecture, the MIHF must be installed at both network elements and mobile nodes. Additionally, an MIIS server must be introduced that has a database with the static attributes of the available access networks in the current EPC domain. We propose the extension of the MIIS server's functionalities so as to support the MIES and MICS services in order to store the reports of the involved entities and enforce policy rules applied by the QoE controller.

A focal point of the described EPC architecture is the PDN-GW, as it interconnects the various access routers such as the Serving-GW for LTE access and the evolved packet data gateway (ePDG) for non-3GPP access. PGW also provides access to external packet data networks like the Internet through the Sgi interface. As a result, it is reasonable to propose that this extended MIH server should be colocated or, at least, should be at the vicinity of the PGW.

3.3. MIH in EPC Architecture. Within the context of this research, MIH is applied as a control mechanism responsible for monitoring and collecting QoE related parameters across different layers and network entities in near real time. Towards this end, MIH primitives including parameters reporting can be utilised. MIH supports different operating modes for parameters reporting. For instance, the `MIH.Link.Get.Parameters` command can be used to obtain the current value of a set of link parameters of a specific link through the `MIH.Link.Parameters.Report` primitive. Additionally, the same primitive can report the status of a set of parameters of a specific link at a predefined regular interval determined by a user configurable timer. Finally, the `MIH.Link.Configure.Thresholds` command can be used to generate a report when a specified parameter crosses

TABLE 1: Summary of QoE related KPIs.

Physical layer	Network layer	Application layer
Access technology	Throughput	Type of service
Signal strength	Delay	Spatial index
	Jitter	Temporal index
	Packet loss	Quality layer
		Quantization parameter

a configured threshold. Subsequently, the reporting can be periodical, event-driven, or explicit depending on the desired approach or the available bandwidth for the monitoring functionality.

The extended MIH server will be the destination of these reports and will store the most recent values of the monitored parameters to the MIIS database that can be shared with the QoE controller. Both static attributes (e.g., the maximum available bandwidth) and dynamic attributes (e.g., the utilised bandwidth) can be stored at the same database. As the MAP is also aware of the characteristics of the video streaming session, the overall set of available monitored parameters will extend from the link layer to the application layer, allowing a more efficient and precise modelling of the perceptual 3D video quality. Figure 5 depicts the basic signalling flow of the periodic reporting scenario. The available links are generating periodic reports of their status and inform their local MIHF through the `Link.Parameters.Report.indication`. Thereafter, the various MIHF report their set of monitoring parameters to MIIS through the `MIH.Link.Parameters.Report.indication` event. Additionally, a similar primitive is generated by the MAP in order to report video related parameters that are available to it.

3.4. QoE Controller

3.4.1. QoE Related Key Performance Indicators. There are three basic categories of QoE parameters related to the physical, network, and application layers that are monitored and collected by MAP and MIH, as shown in Table 1:

- (i) Physical layer related information includes the type of the access technology, the received signal strength, and oscillations of the signal strength. Particularly for the WiFi, this information can easily be collected directly from the access point through queries or broadcast messages.
- (ii) The network QoS parameters are also an important source of information for the QoE management and include the throughput, the end-to-end IP layer delay, jitter, and packet loss. MAP, which is acting as an IP packet filtering function for the IP based video traffic directed to the wireless users, is able to parse the packet header and retrieve the QoE related KPIs.
- (iii) The application parameters monitored involve the type of application (video or audio, although audio is out of the paper's scope), the encoding characteristics, and the video content properties. These are made

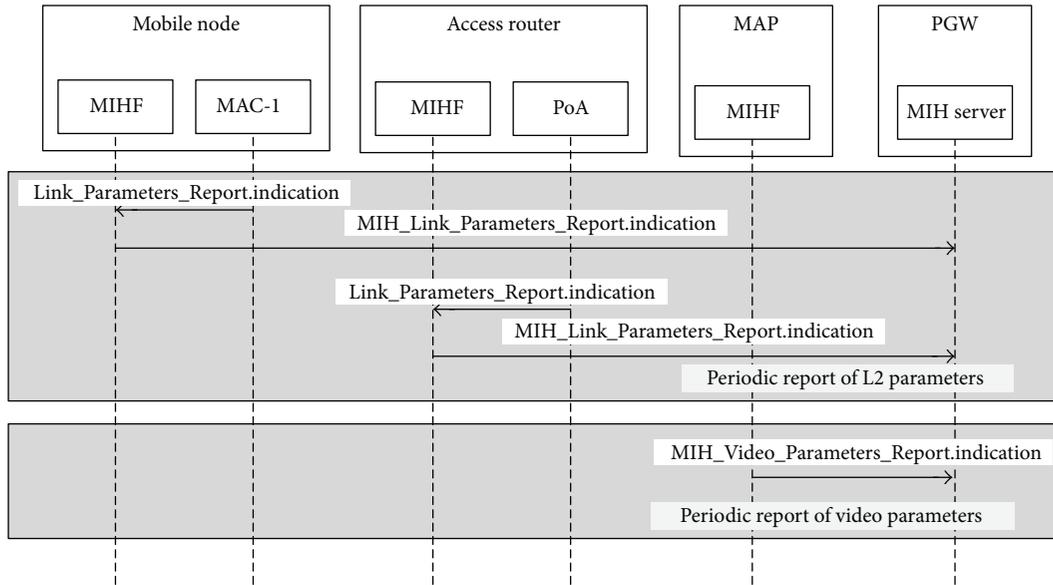


FIGURE 5: Signalling flow of periodic parameters report.

available from the MAP by parsing the RTP header of the received video traffic.

In addition to the parameters of Table 1, a number of additional measurements need to be considered in order to model and monitor the perceived video QoE. Three main groups of measurements are considered: the radio quality, the control plane performance, and the user plane QoS and QoE measurements [31, 32].

The radio quality related measurements will be restricted in just collecting the *number of active user pieces of equipment*. The number of active user devices indicates how many subscribers, on average, use the resources of the cell over a defined period of time and it can be used for traffic and radio resource planning. Moreover, a set of metrics related with the control plane performance are collected, including the *number of connection drops*, where a suddenly occurring exception, which may include loss of signal on the radio interface, can cause interruption of the connection between the user’s equipment and the network. In a scenario that involves real-time video streaming, a loss of an ongoing connection with the content provider will cause extreme deterioration of the user’s experience. In this case, it is mandatory in addition to the *connection drop ratio per cell* to also measure the *connection drop ratio per service*.

Furthermore, user plane QoE related performance indicators utilised by the QoE controller include the *packet jitter* defined as the average of the deviations from the mean network latency. This measurement is based on the arrival time stamp of successive received UDP packets at MAP. The *throughput* is defined as the data volume of IP datagrams transmitted within a defined time period (usually this time period is set to 1 second). In order to determine the IP throughput of each client, MAP parses the IP header and collects and stores the source and destination address of the particular packet. The throughput will be measured at a high

sampling rate (i.e., every 333 ms) for the duration of the ongoing connection.

3.4.2. ML-Based QoE Model. The supervised ML is a learning process based on instances that produce a generalised hypothesis. In turn, this hypothesis could be applied as a mean to forecast future instances [33]. There are a number of common steps distinct in all ML techniques: (a) data set collection, (b) data preprocessing, (c) features creation, (d) algorithm selection, and (e) learning and evaluation using test data. The accuracy of the model improves through the repetition and the adjustment of any step that needs improvement. Supervised ML algorithms could be categorized as follows [34]:

- (i) Logic-based: in this category the most popular algorithms are decision trees, where each node in the tree represents a feature of instances and each branch represents a value that the node can assume. The disadvantage of the algorithms in this category is that they cannot perform efficiently when numerical features are used.
- (ii) Perceptron-based: artificial neural networks are included in this category. Neural networks have been applied to a range of different real-world problems and their accuracy is a function of the number of neurons used as well as the processing cost. However, neural networks may become inefficient when fed with irrelevant features.
- (iii) Statistical: the most well-known statistical learning algorithms are the bit rate Bayesian network and the k-nearest neighbour. The first has the advantage that it requires short computational time for training, but it is considered partial due to the fact that it assumes that it can distinguish between classes using a

single probability distribution. The second algorithm is based on the principle that neighbouring instances have similar properties. Although very simple to use as it requires only the number of nearest neighbours as input, it is unreliable when applied on data sets with irrelevant features.

- (iv) Support Vector Machines (SVM): SVM learning algorithms have been proven to perform better when dealing with multidimension and continuous features as well as when applied to inputs with a nonlinear relationship between them. An example of inputs with nonlinear relationship is the video viewer's QoE and the qualitative metrics gathered from different points of the delivery chain.

Within the context of this study a QoE controller, which is dwelling next to the wireless network provider's core, implements the proposed machine learning-based QoE prediction model. The proposed model could be characterised by fast response time in order to support seamless streaming adaptation, low processing cost, and efficiency. In order to select the best candidate for such a model, three well-known and widely used learning algorithms are investigated and compared in terms of precision and efficiency. The first is bit rate Bayesian classifier [35], the second is a decision tree based on the C4.5 algorithm [36], and the third is a multilayer perceptron network [37].

Naive Bayesian Classifier. A simple method for acquiring knowledge and accurately predicting the class of an instance from a set of training instances, which include the class information, is the Bayesian classifiers. Particularly, the naive Bayesian classifier is a specialized form of Bayesian network that relies on two important simplifying assumptions. Firstly, it assumes that, given the class, the predictive attributes are conditionally independent and, secondly, that the training data set does not contain hidden attributes, which potentially could affect the prediction process.

These assumptions result in a very efficient classification and learning algorithm. In more detail, assuming a random variable C representing the class of an instance, then c is a particular class label. Similarly, X is the vector of random variables that denote the values of the attributes; therefore, x is a particular observed attribute value vector. In order to classify a test case x , the Bayes rule is applied to compute the probability of each class given the vector of observed values for the predictive attributes:

$$p(C = c | X = x) = \frac{p(C = c) p(X = x | C = c)}{p(X = x)}. \quad (1)$$

In (1) the denominator can be calculated given that the event $X = x$ is a conjunction of assigned attribute values (i.e., $X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_k = x_k$) and that these attributes are assumed to be conditionally independent:

$$p(X = x | C = c) = \prod_i p(X_i = x_i | C = c). \quad (2)$$

Evidently, from (2) it is deduced that each numeric attribute is represented as a continuous probability distribution

over the range of the attributes values. Commonly, these probability distributions are assumed to be the normal distribution represented by the mean and standard deviation values. If the attributes are assumed to be continuous, then the probability density function for a normal distribution is

$$p(X = x | C = c) = G(x; \mu_c, \sigma_c) \text{ where,} \\ G(x; \mu_c, \sigma_c) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma}}. \quad (3)$$

Hence, given the class, only the mean of the attribute and its standard deviation are required for the bit rate Bayes classifier to classify the attributes. For the purposes of this study, the naive Bayes classifier will be providing the baseline for the ML-based QoE models in terms of accuracy and precision. This is due to the fact that although it is very fast and robust to outliers and uses evidence from many attributes, it is also characterised by low performance ceiling on large databases and assumes independence of attributes. Moreover, the naive Bayes classifier requires initial knowledge of many probabilities, which results in a significant computational cost.

Logical Decision Tree. By definition a decision tree can be either a leaf node labelled with a class, or a structure containing a test, linked to two or more nodes (or subtrees). Hence, an instance is classified by applying its attribute vector to the tree. The tests are performed into these attributes, reaching one or other leaf, to complete the classification process. One of the most widely used algorithms for constructing decision trees is C4.5 [36]. Although C4.5 is not the most efficient algorithm, it has been selected due to the fact that the produced results would be easily compared with similar research works. There are a number of assumptions commonly applied for C4.5 in order to increase its performance efficiency and classification process:

- (i) When all cases belong to the same class, then the tree is a leaf and is labelled with the particular class.
- (ii) For every attribute, calculate the potential information provided by a test on the attribute, according to the probability of each case having a particular value for the attribute. Additionally, measure the information gain that results from a test on the attribute, using the probabilities of each case with a particular value for the attribute being of a particular class.
- (iii) Depending on the current selection criterion, find the best attribute to create a new branch.

In particular the selection (or splitting) criterion is the normalized information gain. Inherently, C4.5 aims to identify the attribute that possesses the highest information gain and create a splitting decision node. This algorithmic process can be analytically represented with functions (4), assuming that the entropy of the n -dimensional vector of attributes of the sample $H(\vec{y})$ denotes the disorder on the data, while the conditional entropy $H(j | \vec{y})$ is derived from iterating over all possible values of \vec{y} :

$$H(\vec{y}) = - \sum_{j=1}^n \frac{|y_j|}{|\vec{y}|} \log \frac{|y_j|}{|\vec{y}|},$$

$$H(j | \bar{y}) = \frac{|y_j|}{|\bar{y}|} \log \frac{|y_j|}{|\bar{y}|},$$

$$\text{Gain}(\bar{y}, j) = H(\bar{y} - H(j | \bar{y})). \quad (4)$$

The algorithm, ultimately, needs to perform pruning of the resulting tree in order to minimise the classification error caused by the outliers included in the training data set. However, in the case of classifying video perceptual quality with the use of decision trees, the training set contains specialisations (e.g., a low number of very high or very low MOS measurements), and hence the outlier detection and cleansing of the training set needs to be performed beforehand, as described in Section 5.1.

Multilayer Perceptron. A multilayer perceptron network (also known as multilayer feed-forward network) has an input layer of neurons that is responsible for distributing the values in the vector of predictor variable values to the neurons of the hidden layers. In addition to the predictor variables, there is a constant input of 1.0, called the bias, that is fed to each of the hidden layers. The bias is multiplied by a weight and added to the sum going into the neuron. When in a neuron of the hidden layer, the value from each input neuron is multiplied by a weight and the resulting weighted values are added together producing a weighted sum that in turn is fed to a transfer function. The outputs from the transfer function are distributed to the output layer. Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight and the resulting weighted values are added together producing a weighted sum, which is fed into the transfer function. The output values of the transfer function are the outputs of the network. In principle the transfer function could be any function and could also be different for each node of the neural network. In this study the sigmoidal (s-shaped) function has been used in the form of $f(x) = 1/(1 + \exp(-x))$. This is a commonly used function that also is mathematically convenient as it produces the following derivative property: $d\sigma(x)/dx = \sigma(x)(1 - \sigma(x))$.

The training process of the multilayer perceptron is to determine the set of weight values that will result in a close match between the output from the neural network and the actual target values. The algorithm precision depends on the number of neurons in the hidden layer. If an inadequate number of neurons are used, the network will be unable to model complex data, and the resulting fit will be poor. If too many neurons are used, the training time may become excessively long, and, worse, the network may overfit the data. When overfitting occurs, the network will begin to model random noise in the data. In the context of this study several validation experiments have been performed using different number of neurons in the hidden layer. The best accuracy has been achieved by using five neurons in one hidden layer in the network, as shown in Figure 6. The feed-forward multilayer perceptron minimises the error function but it may take time to converge to a solution (i.e., the minimum value) which may be unpredictable due to the error that is added to the weight matrix in each iteration. Therefore, in order to control

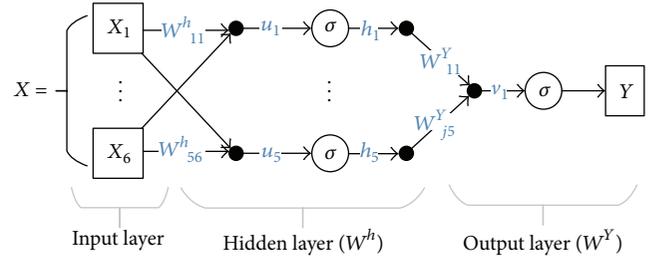


FIGURE 6: Proposed multilayer perceptron model tested and validated.

the convergence rate, the learning rate parameter was set to 0.3 and 350 iterations were performed until the system converged.

4. Experimental Setup

4.1. Capturing and Processing of 3D Video Content. The proposed QoE framework is validated through extensive experiments conducted over the test-bed platform of Figure 8. For the purposes of this study four real-world captured stereo video test sequences (“martial arts,” “music concert,” “panel discussion,” and “report”) with different spatial and temporal indexes have been used in left-right 3D format. The 3D video capturing was performed in accordance with the specifications of [38], in order to provide 3D contents which can be simply classified (genres) for 3D video encoding performances evaluation and comparison, to provide multi-view contents with relevant 3D features from which one can compute disparity maps and synthesise intermediate views to be rendered on different kinds of display, and to allow advances and exhaustive quality assessment benchmarks. Towards this end, an LG Optimus 3D mobile phone was used as stereoscopic capturing device during the shooting session. The cameras were configured as follows:

- (i) Synchronized stereo camera (f2.8, FullHD, 30 fps, Bayer format, 65 mm spaced).
- (ii) Raw formats: bayer format which is then converted to obtain rgb and yuv sequences (+ color correction, noise filtering, mechanical misalignment, and auto-convergence).
- (iii) Encoding: FullHD or lower, side-by-side, or top-bottom with subsampling + MPEG4 AVC.

Snapshots of the test video sequences are shown in Figure 7. A detailed description of the capturing scenarios and content characteristics are described in [38]. All sequences are side-by-side with resolutions 640×720 pixels and 960×1080 pixels per view at 25 frames per second. The H.264/SVC encoding was performed using the encoder provided by *Vanguard Software* tool [39] configured to create two layers (one base layer and one enhancement layer) using MGS quality scalability. The SVC was the favourable choice for the particular experiments, as it allows 3D video content to be delivered over heterogeneous wireless channels with a manageable bit rate. The particular encoder inherently allows



FIGURE 7: Test 3D video sequences used for MOS measurements.

the use of a variety of patent protected error concealment schemes that can compensate for packet losses up to 5%. Each video frame was encapsulated in a single network abstraction layer unit (NALU) with a size of 1400 bytes, which in turn was encapsulated into a single Real-Time Transport Protocol (RTP) packet with a size of 1400 bytes (payload). The generated video packets are delivered through the network using UDP/IP protocol stack. An additional separate channel is responsible for the transmission of the Parameter Sets (PS) to the client through a TCP/IP connection. Moreover, *NetEm* [40] was used for emulating diverse networking conditions, variable delay, and loss in accordance with the described experimental scenarios. Hence, each experiment has been repeated 10 times in order to obtain valid statistical data. Table 2 summarises the encoding and streaming configuration parameters used in all experiments.

Wireless Channel Error Model. In order to model the impact of physical impairments on the QoE, the Rayleigh fading channel of the simulated 802.11g is represented by a two-state Markov model. Each state is characterized by a different bit error rate (BER) in the physical layer that results from the state and transitional probabilities of the Markov model, as in Table 3. The wireless channel quality is characterized by the probabilities of an error in the bad and good state, P_B and P_G , and the probabilities P_{GG} and P_{BB} to remain at the good or the bad state, respectively.

4.2. MAC Layer Load Model. The load of the wireless channel will result in time-outs, which will then cause retransmissions in the data link layer due to the contention-based access that is inherent in IEEE 802.11 protocol. Eventually, since the paper investigates real-time video streaming, such a load in the MAC layer will result in losses that in the application

TABLE 2: Encoding and streaming parameters.

Video sequence	<i>Martial arts</i>	<i>Munich</i>	<i>Panel discussion</i>	<i>Report</i>
Spatial/temporal index	21/17	25/8	32/15	33/3
Number of frames	400 frames			
Intra period	8 frames			
QP used for base-enhancement layers	42–36 & 36–30			
Frame rate	25 frames per second			
Resolution per view	640 × 720 & 960 × 1080 pixels			
Media transmission unit	1500 bytes			

layer will be presented as artefacts and distortion of the perceptual video quality. In order to emulate and control the load, UDP traffic is generated and transmitted to both uplink and downlink channels. The constant-sized UDP packets are generated according to the Poisson distribution three times with mean values of 2 Mbps, 3 Mbps, and 4 Mbps, in each direction, respectively. Evidently, such a setup will eventually double the overall background traffic (i.e., load) over the WiFi access channel.

4.3. IP Delay Variation Model. Moreover, for the purpose of this research and in order to model the delay variations in the IP layer, a constant plus gamma distribution is applied [41], similar to the one depicted in Figure 9. More precisely, the delay variations are modelled as a gamma distribution with a constant scale factor of $\theta = 30$ ms and three different

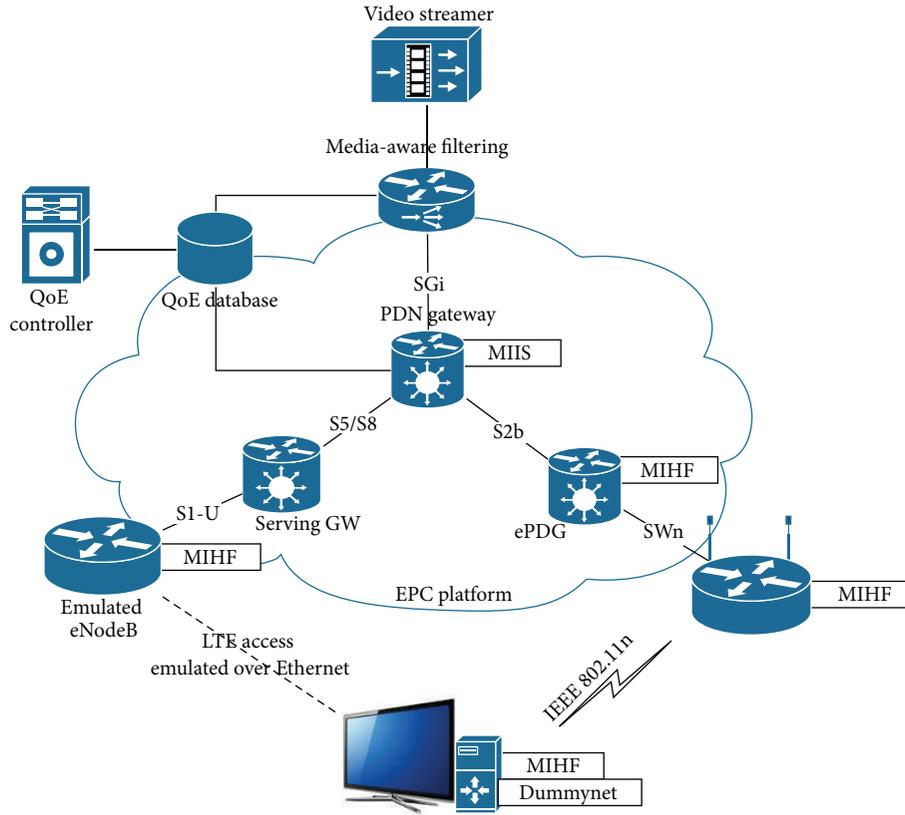


FIGURE 8: Test-bed platform used during the experiments.

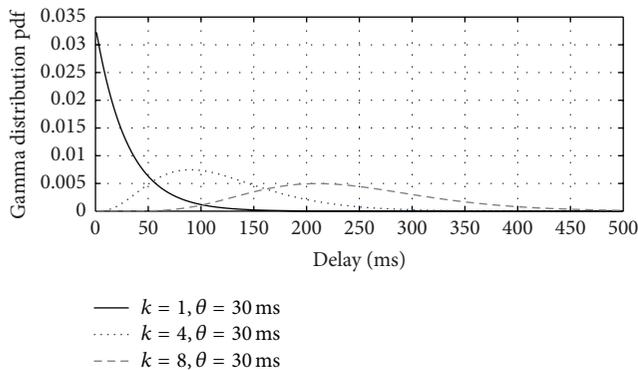


FIGURE 9: Delay variations in IP layer based on gamma distributions.

shape factors in the range of $k = 1, 4, \text{ and } 8$. Apparently, this configuration will result in various mean delays and corresponding variances. During the experiments, three such delay variations schemes were applied, summarised in Table 4, in an effort to control the networking conditions and produce comprehensible outcomes, regarding the visual quality degradation of the decoded video.

5. Subjective Experiments and Analysis

The subjective experiments were conducted in accordance with the absolute category rating (ACR) method as defined in

TABLE 3: BER emulation parameters for the Rayleigh fading wireless channel.

	Bad channel	Medium channel	Good channel
P_G	$1.25e^2$	$1.29e^2$	$1.29e^2$
P_B	$4.13e^{14}$	$1.3e^{13}$	$4.1e^{12}$
P_{GG}	0.996	0.990	0.987
P_{BB}	0.336	0.690	0.740

TABLE 4: IP layer delay variation parameters resulting from gamma distribution.

Shape factor (k)	Mean (ms)	Variance (ms)
1	30	900
4	120	3600
8	140	7200

ITU-T Recommendations [42–45]. The perceived video quality is rated in a scale from 0 [bad] to 5 [excellent] according to the standard. In order to produce reliable and repeatable results, the subjective analysis of the video sequences has been conducted in a controlled testing environment. An LG 32LM620S LED 3D screen with a resolution of 1920×1080 pixels, aspect ratio 16 : 9, peak display luminance 500 cd/m^2 , and contrast ratio of 5000000 : 1 along with passive glasses is used during the experiments to display the stereoscopic video sequences. The viewing distance for the video evaluators is set

to 2 m, in accordance with the subjective assessment recommendations. The measured environmental luminance during measurements is 200 lux, and the wall behind the screen luminance is 20 lux, as recommended by [42]. The subjective evaluation was performed by twenty expert observers in a gender balanced way, who were asked to evaluate the overall visual quality of the displayed video sequences. A training session was included in order for the observers to get familiar with the rating procedure and understand how to recognize artefacts and impairments on the video sequences.

5.1. Statistical Measures. Initially, the collected MOSs were analysed for each subject across the different test conditions using the chi-square test [46], which verified the normality of the majority of MOS distributions. Following the normality verification of the MOS distributions, the technique of outliers' detection and elimination was applied. This technique is based on the detection and removal of scores in the cases where the difference between the mean subject vote and the mean vote for this test case from all other subjects exceeds 15%. The applied outlier detection method determined a maximum of two outliers per test case.

After removing the outliers, the remaining MOSs were statistically analysed in order to assess the perceptual quality of the received video sequences. The average MOS of each set S of tested video sequence scenarios of size K is denoted as $\overline{\text{MOS}}_k$:

$$\overline{\text{MOS}}_k = \frac{\sum_{k=1}^K \sum_{i=1}^N \text{MOS}_{i,k}}{KM}, \quad (5)$$

where $\text{MOS}_{i,k}$ is the Mean Opinion Score of the i th viewer for the k th tested scenario of one of the video sequences and N is number of evaluators of the particular tested scenario. The standard deviation of MOS_k is defined by the square root of the variance:

$$\sigma_k^2 = \frac{\sum_{k=1}^K \left[\sum_{i=1}^N \text{MOS}_{i,k}/N - \overline{\text{MOS}}_k \right]^2}{K-1}. \quad (6)$$

The confidence interval (CI) of the estimated mean MOS values, which indicate the statistical relationship between the mean MOS and the actual MOS, is computed using Student's t -distribution as follows:

$$\text{CI}_j = t \times \left(1 - \frac{\alpha}{2}, N \right) \times \frac{\sigma_j}{\sqrt{N}}, \quad (7)$$

where $t \times (1 - \alpha/2, N)$ is the t -value that corresponds to a two-tailed t -Student distribution with $N - 1$ degrees of freedom and a desired significance level α . In this case, $\alpha = 0.05$, which corresponds to 95% significance, and σ_j is the standard deviation of the MOS of all subjects per test. Furthermore, the degree of asymmetry of the collected MOS around the mean value of its distribution of its measured samples is measured by skewness, while the "peakedness" of the distribution is measured by its kurtosis as follows:

$$\beta = \frac{m_3}{m_2^{3/2}}, \quad (8)$$

$$\gamma = \frac{m_4}{m_2^2},$$

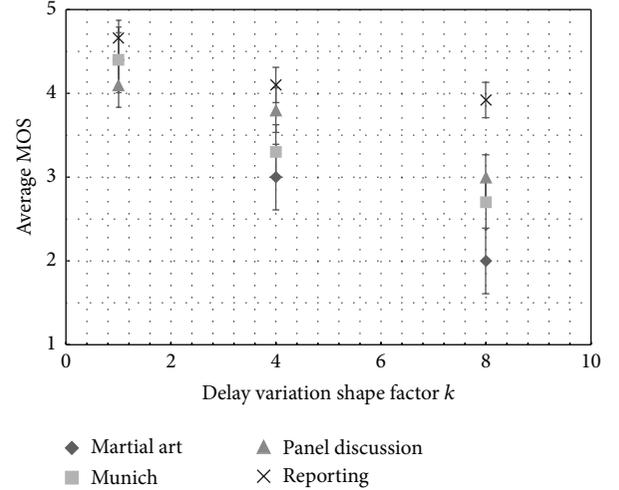


FIGURE 10: Average MOS measured as a result of artefacts due to delay variations.

where m_j is the j th moment about the mean given by

$$m_j = \frac{\sum_{k=1}^K \left[\sum_{i=1}^N \text{MOS}_i/N - \overline{\text{MOS}}_k \right]^j}{K}. \quad (9)$$

Finally the measured MOSs are subjected to an analysis of variance (ANOVA). The ANOVA tests the null hypothesis that the means between two or more groups are equal under the assumption that the sample populations are normally distributed. Although multifactor ANOVA would reveal the complex relationship of BER, MAC layer load, and delay variations with the resulting QoE as in [47], in this study the aim is to determine whether each of the three selected factors, which can be considered as statistically independent, does have a statistically significant impact on the resulting QoE. In this particular case the ANOVA is used in order to test the following null hypothesis:

- (1) The mean MOSs due to the three delay variation groups (i.e., for $k = 1, k = 4$, and $k = 8$) independent of the video sequence, their spatial resolution, and quantization step size are equal.
- (2) The mean MOSs due to the three MAC layer background loads (i.e., 2 Mbps, 3 Mbps, and 4 Mbps) independent of the video sequence, their spatial resolution, and quantization step size are equal.
- (3) The mean MOSs due to the three wireless channel BERs (i.e., Good, Medium, and Bad) independent of the video sequence, their spatial resolution, and quantization step size are equal.

5.2. Analysis of Measured MOS. The averaged MOS measurements of the four test video sequences under the abovementioned experimental conditions are shown in Figures 10, 11, and 12. Each column of the figures represents the average MOS scored by the viewers due to the impact of just one of the three QoS metrics (i.e., BER, MAC load, and delay variation),

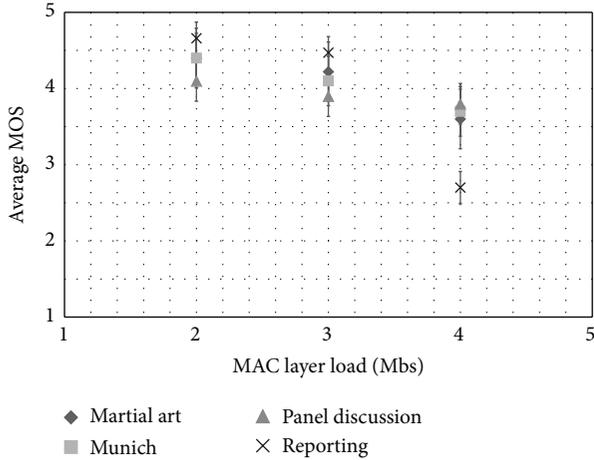


FIGURE 11: Average MOS measured as a result of artefacts due to MAC layer load.

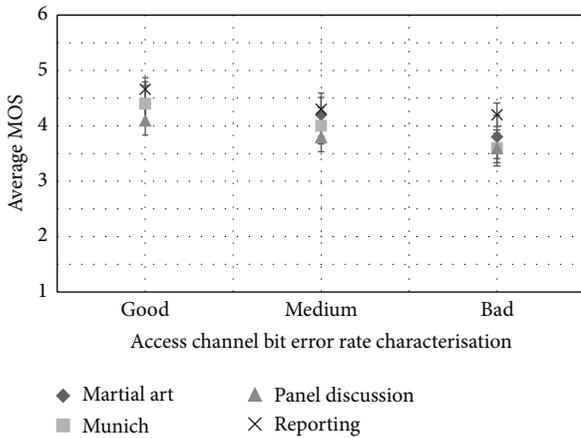


FIGURE 12: Average MOS measured as a result of artefacts due to bit error rate.

while the other two metrics are assigned to their “best” value (i.e., the value that results from the highest level of video delivery QoS). The comparison of the averaged MOSs provides an indication on the impact of the delay variation, load, and BER on the tested sequence. It is evident that the QoS parameters have different effect depending on the video characteristics (i.e., spatial and temporal indexes). Moreover, Table 5 summarises the statistical parameters of MOS per tested video sequence.

In order to get a better understanding of the MOS per tested video sequence the mean, the standard deviation, the variance, the skewness, and the kurtosis are studied. Through variance, the difficulty or ease of the evaluator to assess a parameter as well as the agreement between evaluators can be addressed. Hence, a lower variance may indicate a higher agreement of the overall 3D MOS among viewers. In Table 5 the low variance of the MOS around the mean in all video sequences indicates that viewers agreed on the scores that were assigned to the overall 3D video quality in each test scenario. A slightly higher variance of the “martial arts”

sequence may be due to the high temporal index of the sequence that caused a more erratic behaviour of losses, thus causing more discrepancies among the viewers.

In the context of subjective evaluation, skewness shows the degree of asymmetry of the scores around the MOS value of each distribution of samples for the given parameter (in this case the video sequences). As a normal distribution has a skewness of zero, the results in Table 5 indicate a symmetry of the scores around the MOS for every tested video sequence. Similarly to the study of the variance, the “martial arts” sequence has a slight asymmetry compared to the rest of the sequences due to the fact that the viewers of the sequence approached the maximum and minimum scores more frequently than in the other cases.

The last column of Table 5 presents the kurtosis of the MOS of each video sequence. The kurtosis is the indication of how outlier-prone a distribution is. A normal distribution has a default kurtosis value of 3. The lower the kurtosis, the more prone the distribution. Nevertheless, a kurtosis two times below the standard error of the kurtosis, which is calculated as the square root of 24 divided by N (the number of MOSs) [48], could be considered as normal. In this case, as there are 20 evaluators scoring 108 different scenarios per video sequence, the kurtosis below 40 can be ignored. Therefore, the consistent low kurtosis of Table 5 for all video sequences can be regarded as an indication of lack of outliers, since all evaluators assigned similar scores to the overall 3D video quality.

Three one-way ANOVA methods have been performed in order to evaluate an equivalent number of null hypotheses regarding the three QoS related parameters under study (delay variations, load, and BER). The ANOVA results are summarised in Tables 6, 7, and 8. In more detail, assuming that the grand mean of a set of samples is the total of all the data values divided by the total sample size, then the total variation is comprised as the sum of the squares of the differences of each mean with the grand mean. In ANOVA, the between-group variation and the within-group variation are defined. By comparing the ratio of between-group variance to within-group variance ANOVA determines, if the variance caused by the interaction between the samples is much larger when compared to the variance that appears within each group, then it is because the means are not the same. Additionally, if there are n samples involved with one data value for each sample (the sample mean), then the between-groups variation has $n - 1$ degrees of freedom. Similarly, in the case of within-group variation, each sample has degrees of freedom equal to one less than their sample sizes, and there are n samples; the total degrees of freedom are n less than the total sample size: $df = Nn$, where N is the total size of the samples. The variance due to the differences within individual samples is denoted as MS (Mean Square) within groups. This is the within-group variation divided by its degrees of freedom. The F -test statistic is found by dividing the between-group variance by the within-group variance. The degrees of freedom for the numerator are the degrees of freedom for the between-group variance ($n-1$) and the degrees of freedom for the denominator are the degrees of freedom for the within-group variance ($N - n$). According to the F -test the decision will be to reject the null hypothesis if the F -test statistic from

TABLE 5: Statistical analysis of measured MOS per video test sequence.

Video sequence	Average MOS	Standard dev.	Variance	95% CI	Skewness	Kurtosis
Martial arts	2.390	0.836	0.698	0.391	0.284	2.208
Munich	2.943	0.695	0.483	0.325	0.010	2.288
Panel discussion	3.223	0.570	0.325	0.266	0.173	2.416
Reporting	3.689	0.451	0.203	0.211	-0.024	2.278

TABLE 6: One-way ANOVA for the parameter delay variation.

Source	Sum of squares	Degrees of freedom	MS	F	Probability of F
Between	103.542	2	51.7711	126.95	$4.91566e^{-44}$
Within	174.945	429	0.4078		

TABLE 7: One-way ANOVA for the parameter MAC load.

Source	Sum of squares	Degrees of freedom	MS	F	Probability of F
Between	11.697	2	5.8486	9.4	0.0001
Within	266.79	429	0.62189		

TABLE 8: One-way ANOVA for the parameter BER.

Source	Sum of squares	Degrees of freedom	MS	F	Probability of F
Between	23.852	2	11.9262	20.09	$4.55639e^{-09}$
Within	254.635	429	0.5936		

Tables 6, 7, and 8 is greater than the F -critical value with $k - 1$ numerator and $N - k$ denominator degrees of freedom.

By studying Tables 6, 7, and 8, it can be affirmed that all three null hypotheses, as defined previously, can be rejected; hence, there is significant statistical difference among the groups with respect to delay variations (i.e., the different shape factors k of the gamma distribution have a significant effect on the resulting 3D video MOS), MAC layer load (i.e., the three different mean values of the Poisson distributed background traffic affect significantly the 3D MOS), and the BER (i.e., the three different wireless channel conditions alter the 3D MOS significantly). The above conclusions are derived by the extremely low probability p which is much lower than the significance level of 0.05 and the F -statistic which in turn indicates that one or two means significantly differ between each other.

Therefore, one-way ANOVA affirms that the selection of the three QoS parameters (delay variation, load, and BER) as well as the values assigned to these parameters for the purposes of the subjective evaluation of 3D MOS and subsequently the definition, training, and validation of a machine learning QoE model, based on these QoS parameters, is correct.

6. ML QoE Models Comparison

The training and the analysis of the results have been performed using Weka software tool [49]. The output of the naive Bayes classification is shown in Table 9, where for every attribute of the data set the mean and standard deviation of the Gaussian distribution are shown. Furthermore, the

decision tree of the C4.5 ML algorithm that is implemented as J48 in Weka is illustrated in Figure 13. In accordance with the results of the MOS comparison in the previous section, the jitter is the most important parameter; hence, it is located in the route of the tree (i.e., first splitting node).

All models are assessed in terms of predictive performance using the 10-fold cross-validation method. The confusion matrix of each of the three models is shown in Table 10. It can be seen that the naive Bayesian classifier has classified wrongly 42 instances of MOS class 2, 43 instances of MOS class 3, and all MOS class 5 instances. This is expected since the naive Bayes classifier was selected to act as the baseline of the ML performance tests. The confusion matrices of multilayer perceptron and C4.5 decision tree confirm that both methods perform similarly and better than the Bayesian classifier.

Nevertheless, the models accuracy (number of correctly classified instances) cannot be used for assessing the usefulness of classification models built using unbalanced datasets, as in this case. For this purpose the “Kappa statistic” is used, which is a generic term for several similar measures of agreement used with categorical data. Typically it is used in assessing the degree to which two or more video viewers in this case, examining the same video sequence, agree when it comes to assigning the MOSs. Similarly to the correlation coefficient, its value is zero when there is no agreement on the scores (purely random coincidences of rates) and is one when there is complete agreement. The “Kappa statistic” of the three models is shown in Table 11.

Additionally, the three ML algorithms considered in this work are compared with each other in terms of accuracy,

TABLE 9: Output of naive Bayes classifier.

Attributes		Class (MOS)				
		1	2	3	4	5
SI	Mean	20.21	21.71	22.28	22.67	24
	Std. dev.	0.89	1.97	1.98	1.88	0.6667
TI	Mean	17.53	14.16	12.87	12	9
	Std. dev.	2.01	4.45	4.45	4.24	1.5
QP ₁	Mean	39.15	38.82	39.55	38.14	36
	Std. dev.	2.99	2.99	2.94	2.87	1
QP ₂	Mean	33.15	32.82	33.55	32.14	30
	Std. dev.	2.99	2.99	2.94	2.87	1
Resolution	Mean	0.57	0.51	0.46	0.5	1
	Std. dev.	0.49	0.49	0.49	0.5	0.16
Jitter	Mean	7	5.61	2.68	0.25	0
	Std. dev.	0.58	2.00	2.17	0.90	0.5833
Load	Mean	3.42	3.13	2.98	2.64	2
	Std. dev.	0.67	0.82	0.81	0.71	0.16
Channel	Mean	0.52	0.85	1.06	1.28	2
	Std. dev.	0.67	0.77	0.83	0.76	0.16

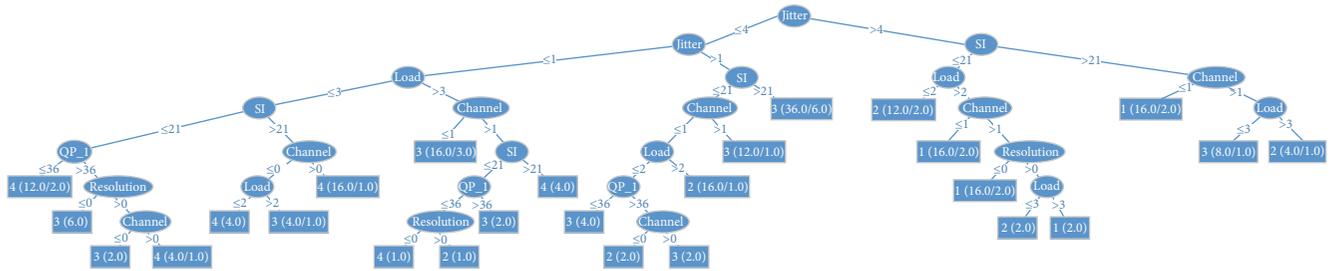


FIGURE 13: C4.5 based classification tree of the proposed model.

precision (p), and recall bit rate. The precision of the algorithm, which is also known as reproducibility or repeatability, denotes the degree to which repeated measurements (in this case of MOS) under unchanged conditions show the same results. Recall is defined as the number of relevant instances retrieved by a search divided by the total number of existing relevant instances. In addition to the two metrics, the algorithms are also evaluated in terms of F -measure (F), which considers both the precision (p) and the recall (r) of the test to compute the score. Let $i \in [1, 2, 3, 4, 5]$ represent the class, which in this case is derived from the MOSs. Then true positive TP_i is the number of the instances correctly classified, false positive FP_i is the number of the instances that belong to the class bit rate but have not been classified there, and false negative FN_i is the number of the instances that do not belong in class bit rate but have been classified there. Thus, (p), (r), and (F) are derived as follows:

$$p = \frac{TP_i}{TP_i + FP_i},$$

$$r = \frac{TP_i}{TP_i + FN_i},$$

$$F = \frac{2pr}{p + r}.$$

(10)

Finally, the accuracy of the models is also deduced by the area under the Receiver Operating Characteristic (ROC) curve. An area of one represents a perfect model, while an area of 0.5 means lack of any statistical dependencies. These comparison results are also shown in Table 11.

It is evident that for the particular data set, with the multiple classes and the nonlinear relationship among the attributes, the best performance is achieved by the multilayer perceptron. Nevertheless, C4.5 is a very good alternative that performs MOS classification with acceptable accuracy and with less processing effort. Moreover, since decision trees return results that have a physical significance, thus the classification can be easily interpreted and used as a decision

TABLE 10: Confusion matrices of the three ML models.

a	b	c	d	e	← classified as
Naive Bayesian confusion matrix					
2	17	0	0	0	$a = 1$
2	39	39	1	0	$b = 2$
0	11	134	32	0	$c = 3$
0	0	40	109	0	$d = 4$
0	0	0	6	0	$e = 5$
Multilayer perceptron confusion matrix					
11	8	0	0	0	$a = 1$
6	63	12	0	0	$b = 2$
0	13	141	23	0	$c = 3$
0	0	21	128	0	$d = 4$
0	0	0	6	0	$e = 5$
C4.5 confusion matrix					
12	7	0	0	0	$a = 1$
6	63	12	0	0	$b = 2$
0	14	137	26	0	$c = 3$
0	1	25	121	2	$d = 4$
0	0	0	5	1	$e = 5$

TABLE 11: Performance comparison of ML algorithms.

	Naive Bayesian	Multilayer perceptron	C4.5
Kappa statistic	0.473	0.693	0.664
Precision (p)	0.643	0.782	0.770
Recall (r)	0.657	0.794	0.773
F -measure (F)	0.641	0.788	0.771
Area under ROC	0.851	0.912	0.886

rule; C4.5 would be preferable as opposed to the neural network, as a decision engine during real-time QoE control.

7. Conclusion and Future Steps

Evidently, a new wave of IP convergence with 3GPP's LTE EPC in its heart is being fuelled by real-time multimedia applications. As new network architecture paradigms are constantly evolving, there is an ongoing effort by content providers and operators to provide and maintain premium content access with optimised QoE. This paper presents a novel media-aware framework for measuring and modelling the 3D video user experience, designed to be closely integrated with LTE EPC architecture. The proposed solution involves a synergy between a Media-Aware Proxy networking element and the IEEE 802.21 MIH protocol. The former is responsible for parsing RTP packets to collect QoE related KPIs and adapting the video streams to the current networking conditions in real time. The latter is being enhanced to act as control mechanism for providing the necessary signalling to collect KPIs across different layers and network elements. Moreover, MIH provides the central database, which stores the collected QoE related KPIs and allows a third component

named QoE controller to train this dataset and model the perceived 3D video quality. Three machine learning classification algorithms for modelling QoE due to network related impairments have been investigated. Opposite to previous studies, the QoE model is a function of parameters collected from the application, the MAC, and physical layers. Hence, instead of determining the 3D QoE by measuring only the end-to-end packet loss, this paper considers a set of quality of service (QoS) parameters (i.e., bit error rate from physical layer, MAC layer load, and network layer delay/jitter), which account for the overall number of lost packets. A detailed statistical analysis of the measured 3D MOS indicated that the predominant factor of QoE degradation is the IP layer jitter. Therefore, the proposed ML scheme aims to determine more accurately the impact that different layer impairments have on the perceived 3D video experience. A comparison of the three ML schemes under study in terms of precision and accuracy indicated that the multilayer perceptron has the best performance, closely followed by the C4.5 decision tree.

This is an ongoing research. Building upon the results presented in this paper, the following steps involve the design and implementation of a QoE management mechanism that will optimise the perceived video QoE by instructing different network elements of the EPC architecture and the MAP to either adapt the delivered video traffic to the network conditions or enforce with the aid of MIH a content-aware handover to a less loaded neighbouring channel or access network. Moreover, a closer integration with the current and future LTE EPC implementations is currently under study. The aim is to reconfigure the proposed framework to incorporate the LTE policy and charging control (PCC) techniques, in an effort to upgrade it into a content-aware traffic mechanism extended over all network operator deployments.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

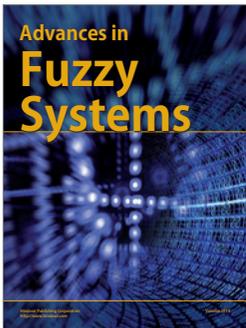
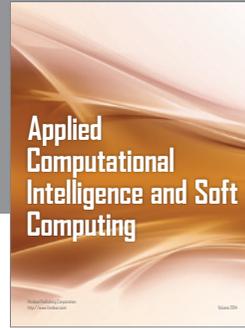
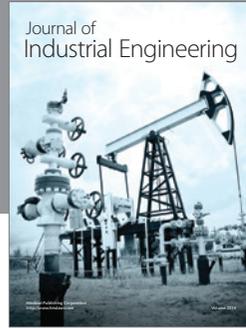
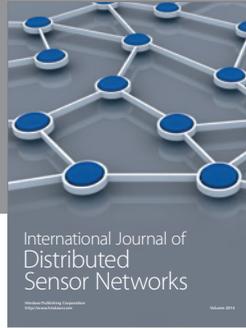
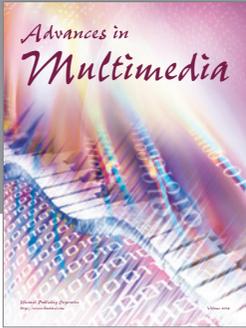
This work was supported by the SIEMENS "Excellence" Program by the State Scholarship Foundation (IKY), Greece.

References

- [1] A. Lykourgiotis, K. Birkos, T. Dagiuklas et al., "Hybrid broadcast and broadband networks convergence for immersive TV applications," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 62–69, 2014.
- [2] G.-M. Su, Y.-C. Lai, A. Kwasinski, and H. Wang, "3D video communications: challenges and opportunities," *International Journal of Communication Systems*, vol. 24, no. 10, pp. 1261–1281, 2011.
- [3] C. Tselios, I. Politis, K. Birkos, T. Dagiuklas, and S. Kotsopoulos, "Cloud for multimedia applications and services over heterogeneous networks ensuring QoE," in *Proceedings of the IEEE 18th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD '13)*, pp. 94–98, Berlin, Germany, September 2013.

- [4] K. Birkos, I. Politis, A. Lykourgiotis et al., "Towards 3D video delivery over heterogeneous networks: the ROMEO approach," in *Proceedings of the International Conference on Telecommunications and Multimedia (TEMU '12)*, pp. 60–65, IEEE, Chania, Greece, August 2012.
- [5] ITU-T and ISO/IEC, "ITU-T recommendation H.264 (2010): advanced video coding for generic audiovisual services," Information Technology Coding of Audiovisual Objects Part 10: Advanced Video Coding ISO/IEC 14496-10:2010, 2010.
- [6] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, 2011.
- [7] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [8] 3GPP, "Technical specification group services and system aspects," 3GPP TS 23.002 V13.3.0 (2015-09), Network Architecture, 2015.
- [9] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.
- [10] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, "Joint draft 8 of multi-view video coding," Tech. Rep. JVT-AB204, Joint Video Team (JVT), Hannover, Germany, 2008.
- [11] G. J. Sullivan, A. M. Tourapis, T. Yamakage, and C. S. Lim, "Draft AVC amendment text to specify constrained baseline profile, stereo high profile, and frame packing SEI message," Joint Video Team (JVT) Document JVT-AE204, Joint Video Team (JVT), London, UK, 2009.
- [12] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461–1473, 2007.
- [13] I. Politisa, L. Dounis, C. Tselios, A. Kordelas, T. Dagiuklas, and A. Papadakis, "A model of network related QoE for 3D video," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps '12)*, pp. 1335–1340, Anaheim, Calif, USA, December 2012.
- [14] C. Tselios, C. Mysirlidis, I. Politis, T. Dagiuklas, and S. Kotsopoulos, "On quality evaluation of 3D video using Colourplus-Depth and MDC," in *Proceedings of the IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD '14)*, pp. 66–69, IEEE, Athens, Greece, December 2014.
- [15] Nokia-Siemens Networks, "QoS-enabled IP networks for assured multiplay QoE," Technical White Paper, 2011.
- [16] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "Quadrant of euphoria: a crowdsourcing platform for QoE assessment," *IEEE Network*, vol. 24, no. 2, pp. 28–35, 2010.
- [17] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 339–350, 2013.
- [18] K. U. R. Laghari and K. Connelly, "Toward total quality of experience: a QoE model in a communication ecosystem," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 58–65, 2012.
- [19] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. De Veciana, "Video quality assessment on mobile devices: subjective, behavioral and objective studies," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, 2012.
- [20] V. De Silva, H. K. Arachchi, E. Ekmekcioglu et al., "Psychophysical limits of interocular blur suppression and its application to asymmetric stereoscopic video delivery," in *Proceedings of the 19th International Packet Video Workshop (PV '12)*, pp. 184–189, IEEE, Munich, Germany, May 2012.
- [21] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access," 3GPP TS 23.401, 2016.
- [22] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Network Architecture," Tech. Rep. 3GPP TS 23.002 V12.5.0, 2014.
- [23] IETF, "Diameter base protocol," Tech. Rep. RFC 6733, IETF, 2012.
- [24] ETSI TS 123 228 V12.9.0, "IP Multimedia Subsystem (IMS), stage 2," 3GPP TS 23.228 version 12.9.0 Release 12, 2015.
- [25] Fraunhofer FOKUS OpenEPC Toolkit, <http://www.openepc.net/>.
- [26] IEEE 802.21 Working Group, "IEEE standard for local and metropolitan area networks part 21: media independent handover," IEEE Standard 802, 2008.
- [27] A. Kordelas, I. Politis, A. Lykourgiotis, T. Dagiuklas, and S. Kotsopoulos, "An media aware platform for real-time stereoscopic video streaming adaptation," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '13)*, pp. 687–691, Budapest, Hungary, June 2013.
- [28] IETF, "RTP payload format for H.264 video," IETF Draft RFC 3984, IETF, 2010, draft-ietf-avt-rtp-rfc3984bis-12.
- [29] IETF Draft Audio/Video Transport WG, "RTP Payload Format for SVC Video," draft-ietf-avt-rtp-svc-27, 2011.
- [30] 3GPP, "Architecture enhancements for non-3GPP accesses," 3GPP TS 23.402, 2015.
- [31] H. Ekström, "QoS control in the 3GPP evolved packet system," *IEEE Communications Magazine*, vol. 47, no. 2, pp. 76–83, 2009.
- [32] H. Luo, S. Ci, D. Wu, J. Wu, and H. Tang, "Quality-driven cross-layer optimized video delivery over LTE," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 102–109, 2010.
- [33] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra Sánchez, "Predicting quality of experience in multimedia streaming," in *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia (MoMM '09)*, pp. 52–59, Kuala Lumpur, Malaysia, December 2009.
- [34] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
- [35] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI '95)*, pp. 338–345, Montreal, Canada, August 1995.
- [36] J. R. Quinlan, *C4.5: Programs for Machine Learning*, vol. 1, Morgan Kaufmann, Boston, Mass, USA, 1993.
- [37] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14–15, pp. 2627–2636, 1998.
- [38] EU-FP7-ICT-ROMEO, "D3.1 Report on 3D media capture and content preparation," May 2012, <http://www.ict-romeo.eu/>.
- [39] Vanguard Software Solutions H.264/SVC SDK, <http://www.vanguardsw.com/>.
- [40] S. Hemminger, "Network emulation with NetEm. , Linux conf au," in *Proceedings of the Australia's National Linux Conference (linux.conf.au '05)*, Canberra, Australia, April 2005.

- [41] J. C. Bolot, “End-to-end packet delay and loss behavior in the Internet,” in *Proceedings of the Conference on Communications Architectures, Protocols and Applications (SIGCOMM '93)*, pp. 289–298, 1993.
- [42] ITU Radio Communication Sector, “Methodology for the subjective assessment of the quality of television pictures,” Tech. Rep. ITU-R BT.500-11, 2002.
- [43] ITU Radio Communication Sector, “Subjective video quality assessment methods for multimedia applications,” Tech. Rep. ITU-T, Rec. P.910, 1999.
- [44] ITU, “Subjective audiovisual quality assessment methods for multimedia applications,” ITU-T Recommendation P.911, International Telecommunication Union, Geneva, Switzerland, 1998.
- [45] ITU-T BT.2021, *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, International Telecommunication Union, Geneva, Switzerland, 2012.
- [46] DTIC, *Chi-Square Tests*, Defense Technical Information Center, 1976.
- [47] A. Khan, L. Sun, and E. Ifeachor, “QoE prediction model and its application in video quality adaptation over UMTS networks,” *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 431–442, 2012.
- [48] B. G. Tabachnick, L. S. Fidell, and S. J. Osterlind, *Using Multivariate Statistics*, 2001.
- [49] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

