

5G-QoE: QoE Modelling for Ultra-HD Video Streaming in 5G Networks

James Nightingale, *Member IEEE*, Pablo Salva-Garcia, Jose Alcaraz Calero, *Senior Member IEEE* and Qi Wang

Abstract— Traffic on future Fifth-Generation (5G) mobile networks is predicted to be dominated by challenging video applications such as mobile broadcasting, remote surgery and augmented reality, demanding real-time and ultra-high quality delivery. Two of the main expectations of 5G networks are that they will be able to handle Ultra High Definition (UHD) video streaming and that they will deliver services that meet the requirements of the end user’s perceived quality by adopting Quality of Experience (QoE) aware network management approaches. This paper proposes a 5G-QoE framework to address the QoE modelling for UHD video flows in 5G networks. Particularly, it focuses on providing a QoE prediction model that is both sufficiently accurate and of low enough complexity to be employed as a continuous real-time indicator of the ‘health’ of video applications flows at the scale required in future 5G networks. The model has been developed and implemented as part of the EU 5G PPP SELFNET autonomic management framework, where it provides a primary indicator of the likely perceptual quality of UHD video application flows traversing a realistic multi-tenanted 5G mobile edge network testbed. The proposed 5G-QoE framework has been implemented in the 5G testbed, and the high accuracy of QoE prediction has been validated through comparing the predicted QoE values with not only subjective testing results but also empirical measurements in the testbed. As such, 5G-QoE would enable a holistic video flow self-optimisation system employing the cutting-edge Scalable H.265 video encoding to transmit UHD video applications in a QoE-aware manner.

Index Terms— QoE, 5G networks, video streaming, UHD.

I. INTRODUCTION

VIDEO applications currently account for 73% of all IP based Internet traffic [1] and are predicted to consume 82% by 2021. Over the same period traffic from mobile devices is set to rise from 7% of all traffic to 17%, increasing at twice the rate of fixed IP traffic. Also as Fifth Generation (5G) mobile networks [2],[3] enter service, expected higher bandwidths, lower end to end delays and improved reliability; are likely to increase demand for mobile video consumption.

Similarly, new video compression standards such as High

(Corresponding author: Qi Wang)

This work has been funded in part by the European Commission Horizon 2020 5G PPP Programme under grant agreement H2020-ICT-2014-2/671672 - SELFNET (Self-Organised Network Management in Virtualised Software Defined networks). This work was additionally funded by the UWS 5G Video Lab project.

All authors are currently with the Artificial Intelligence, Visual Communications and Networks (AVCN) research centre at the University of the West of Scotland, Paisley, United Kingdom. (email: firstname.lastname@uws.ac.uk).

Efficiency Video Coding (H.265/HEVC) [4][5] and the availability of Ultra-High-Definition (UHD) portable consumer devices may further fuel growth in mobile video traffic. Some portable devices already have screen resolutions of 4K with 8K possible by the early 2020’s. An 8K laptop screen, using version 1.4 of the embedded DisplayPort standard (eDP) [6] has already been demonstrated Japan Display [7])

These two technological advances will provide the infrastructure for ‘anywhere anytime’ access to real time broadcast media and possibly inspire new classes of video services, again increasing the video related load on mobile networks. Despite anticipated improvements in Quality of Service (QoS) and resilience [8] in 5G networks, enormous volumes of video traffic will continue to pose significant challenges for network operators. Recently, the network quality focus has changed from a network provider’s QoS perspective to the less easily quantified end user’s Quality of Experience (QoE) viewpoint.

In this context, the EU 5G PPP SELFNET project [9][10] has proposed a QoE-aware Self-Optimisation Use Case for UHD video flows using the Scalable H.265 video coding standard. The key enabler in this use case is a QoE prediction model for Scalable H.265 encoded UHD video flows in 5G infrastructures. There are a number of technical challenges to achieve this enabler, as explained below.

Firstly, finding a reliable, accurate, scalable and robust QoE prediction model for streamed video over mobile networks is an unresolved and very challenging task. The specific set of challenges investigated in this paper cover the immensely important area of delivering UHD video to demanding users in 5G mobile networks. These include significantly increased bandwidth, the predicted growth video streaming traffic and subjective factors such user expectations 5G networks.

Secondly, current QoE models including those promoted by standardisation bodies [11], do not focus on 5G networks where additional challenges such as virtualisation, mobility and multi-tenancy requirements exist.

Thirdly, although video encoder type is a significant factor in QoE modelling [12], existing QoE models usually only consider single layer video encoders mostly for the H.264 Advanced Video Coding standard (H.264/AVC) [13] or in a small number of cases the latest H.265 standard [4][5].

To address the above challenges, this paper investigates QoE prediction of UHD video, encoded using the scalable extension to the H.265 standard (SHVC) [14] over 5G networks. By focusing on fast and efficient prediction of QoE

from 5G network congestion indicators, it can predict the QoE of the whole scalable video stream and estimate the QoE achieved by dropping a layer (or layers) from a scalable H.265 video stream. This model is one of the components of the SELFNET autonomic 5G network management system [10]

This work addresses real time, RTP based video streaming often used for video conferencing, video chat and video surveillance applications rather the Dynamic Streaming over HTTP (DASH) based streaming used in Content Delivery Networks (CDNs) such as Netflix [15] or Hulu [16] where a number of pre-recorded and pre-encoded representations of a video stream serve different client types and network conditions.

The model was developed and evaluated through subjective evaluation experiments using over 50 human subjects. Validation compared the results of further subjective evaluations with those predicted by the model. Empirical results show that, for a range of different content types, the predictions of QoE produced by the model closely tracked the subjective opinions of the test subjects.

In summary, this paper will highlight the following novel contributions:

- A 5G-QoE framework comprising essential building blocks to enable the chain of sensing/monitoring, aggregation, QoE modelling and QoE prediction;
- A low-complexity QoE estimation and prediction scheme that is practical to be deployed in real-world networking environment with real-time processing requirements;
- A 5G-aware QoE system that is capable of extracting video metadata and flow QoS metrics to enable the QoE modelling for video flows over a multi-tenancy 5G infrastructure;
- A UHD capable, Scalable H.265 (and H.265) aware QoE system ready for the emerging next generation mainstream video applications in 5G and Internet.

The rest of the paper is organised as follows, in Section 2 the state of the art in QoE modelling for streamed video, scalable video codecs and, where relevant in this context, advances towards autonomic functionality in 5G networks. Section 3 provides an insight into the QoE-driven, self-optimising features of the SELFNET 5G network management architecture, whilst Section 4 explains the methodology used and the subjective testing experiments undertaken. In Section 5, the QoE prediction model is developed and the results of validation experiments presented. Finally, Section 6 concludes the paper.

II. RELATED WORK

This section reviews existing QoE modelling techniques and highlights key technologies relevant to this work.

A. QoE modelling Approaches

Existing QoE assessment and modelling for video can be divided into two broad categories, subjective or objective.

Irrespective of which modelling technique has been employed, all QoE models, through some function or mapping, provide a prediction of the perceived subjective quality of a video under a given set of circumstances. The metric used in these models is normally predicted Mean Opinion Score (MOS). QoE prediction models are commonly validated by comparing the outputs of the model with the results of subjective (from human subjects) evaluations of quality. Where models target a networking environment, they may be further validated experimentally using a network simulator or a testbed. As there are several recent comprehensive survey papers (e.g., [17], [18] and [19]) in this domain, this subsection only summarises the technical approaches that are most relevant to this paper.

1) Subjective QoE Assessment and Models

Subjective QoE assessment methods employ organised sessions of end users who view video content and rate the visual quality using a Mean Opinion Score (MOS) metric. ITU-T recommendations [20] for subjective quality evaluation follow strict t setup and testing conditions. MOS scores are considered to reliably reflect the quality perceived by the Human Vision System (HVS) and therefore, can also be used to validate an objective QoE model. Nevertheless, subjective QoE tests are time-consuming, labour-intensive, expensive and do not scale. Additionally, subjective testing does not provide an instantaneous QoE metric suitable for real-time video assessment or prediction.

Subjective video quality models attempt to leverage insights into HVS through psychological or psychophysiological factors such as user expectations of a service, service type, age, mood and time of day to predict how a user will perceive the quality of a particular video. For instance, Reiter et al [21] have shown that age, sex and socio-economic status are all factors influencing QoE, while Kara et al. [22] claim that economic context such as the brand perception of viewing device (in their case a smartphone) and the price, if any, paid to view the content were significant factors. However, such factors are also difficult to manage and correlate in a unified model for efficient, real-time systems.

2) Objective QoE Models

In light of the drawbacks of subjective QoE assessment and modelling, objective QoE modelling has gained significant popularity over the years. Some models directly map an objective measurement of video quality such as the well-known Peak Signal to Noise ratio (PSNR) and Structural Similarity Index Matrix (SSIM) metrics directly to a prediction of user perceived quality. However, these metrics are often criticised for either requiring full (or reduced) reference comparisons to the original video frames or for being unreliable.

Consequently, practical, no reference QoE modelling is highly desirable. Parametric QoE models, which derive a predicted MOS from a model that is a function of some number of objectively measured parameters, are now the most commonly used objective method of modelling QoE [47]. These parameters have often included Quality of Service (QoS) metrics such as bandwidth, delay, packet loss, bit error

rate etc. In some cases the parameters used have also considered the nature of the video stream being transmitted such as content type, resolution, frame rate etc. QoS to QoE mapping, by exploring and establishing a relationship between QoS metrics and QoE for specific use cases is a primary way to achieve such objective metrics. For instance, the H.264/AVC-encoded 3D video model proposed by Alreshoodi et al. [23], maps QoS parameters from both the video encoding layer (content type, spatial resolution and quantization parameter) and the network layer (packet loss rate and mean burst length). Their model was developed using fuzzy logic inference systems, and may have significant system complexity and computational power requirements. Seyedehbrahimi et al. [24] developed a QoE metric called Pause Intensity (PI) for TCP-based video streaming, in which the PI is determined from video playout rate and network throughput. PI is shown to be the ratio of the rate difference ($\lambda - \eta$) to the playout rate λ . The model was validated through simulations using video sequences encoded with H.264. Another QoS/QoE function was proposed by Hsu and Lo [25] for cloud-based multicast video streaming using a simulated platform. Finally, Khan [26] et al. designed a QoE metric for H.264 video in 3G networks simulated by the ns2 simulator.

Compared with these existing studies, this paper proposes a new objective QoE model for UHD video streaming encoded using the latest standard Scalable H.265. The modelling methodology has leveraged subjective QoE assessment information and has been validated using both subjective and objective approaches and empirically validated in a realistic 5G testbed.

B. 5G UHD Networking, QoE and H.265 Standards

1) 5G UHD Networking and QoE Modelling Requirements

Driven by ever-growing user requirements and expectations, research on future 5G networks has gained global momentum. In Europe, the SELFNET project [9], one 19 EU 5G PPP Phase 1 projects [27], focuses on cognitive network management. A primary use case of SELFNET is self-optimisation of UHD video streaming in 5G hotspots in venues such as airports, and stadiums. The aim is to deliver high-quality (potentially UHD) video streams, from a network media server to multiple users in a 5G hotspot, with sustained levels of user QoE. The main task being to maintain QoE in the face of network congestion and decreasing available bandwidth caused by many users concurrently streaming video.

QoE awareness and prediction are the key enablers in SELFNET's self-optimisation scheme to mitigate the impact of congestion on users. The design and prototype aims to deliver a complete operational chain including practical sensors that gather network metrics and video metadata needed to build the QoE model. In contrast, most existing related work either just assumes the availability of the required metrics/metadata or use simulated values in their modelling. However, it is not trivial to obtain the required metrics/metadata, especially at run-time, for real-time QoE prediction.

Additionally, other 5G requirements [28][29] need to be addressed. New virtualisation, cloud computing and softwareisation technologies (e.g. Software-Defined Networking (SDN) and Network Function Virtualisation (NFV)) will reduce capital and operational expenditure in 5G networks. Multi-tenancy is required as a built-in feature to allow the infrastructure to be shared by two or more operators, and mobility support is fundamental in 5G mobile networks (e.g., through the General Packet Radio Service (GPRS) Tunnelling Protocol (GTP) based on the evolution of the Fourth Generation (4G) Long-Term Evolution (LTE) networks). All these requirements have implications on the design and implementation of the 5G-QoE system.

To the best of our knowledge, no existing 5G video related projects have considered all the above requirements. For instance, the UHD-on-5G project [30] focusses on Information-Centric Networking (ICN) and SDN technologies for efficient 5G UHD video streaming with no consideration of QoE. Kourtis et al. [31] proposed a Video Quality Assessment (VQA) method based on SSIM to address quality degradation introduced by a bottleneck of the small cell backhaul. This work uses a reduced reference, rather than our preferred no reference approach, and does not consider the latest codecs, UHD or multi-tenancy. Ge et al. [32] address 4K UHD video delivery in 5G networks and focus on an adaptive video prefetching scheme deployed at the edge of the network to help improve QoE. However, no QoE modelling is reported apart from using video buffer status as an indicator of a video session's QoE. In addition, no codec or multi-tenancy support is mentioned.

2) H.265 and Scalable H.265 Video Coding Standards, and QoE Studies

The H.265/HEVC standard introduced in 2013 reduces bandwidth requirements, compared with H.264/AVC by up to 50% with no perceptual loss of quality [35]. Recent speed improvements in H.265 (e.g., in [36]), demonstrate its potential to replace H.264 codecs in 5G networking, while interest on the impact of H.265 on network has also gained traction [37].

Despite the need, acknowledged by standardisation bodies (3GPP [33]), and industry alliances (NGMN [34]), for new high compression codecs such as H.265 to reduce bandwidth requirements in 5G networks; the vast majority of related work employs H.264 or older video coding standards. H.265 codecs will help mitigate the vastly increased bandwidths needed by UHD video with potential spatial resolutions of up to 8K and frame rates of up to 300 frames per second (fps).

The current (4th version) of the H.265 standard [5], fully supports scalable video encoding using a range of scalability options including spatial, quality, bit depth and colour gamut. Temporal scalability is native to the H.265 standard itself. Scalable video encoders allow a video stream to be adapted within the network by dropping layers from the stream [38]. Fig. 1 shows the enhancement layer on a scalable stream (Enh1) being dropped at a media adaptation element while the base layer is delivered.

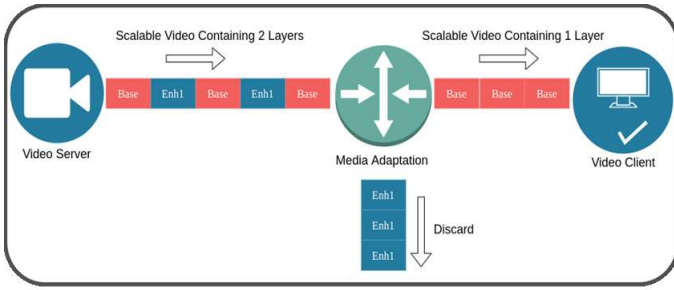


Fig. 1 Video stream adaptation using scalable video codecs.

Existing publications on the scalable extension to H.265 in the context of UHD video, have only made subjective comparisons with previous codecs [39] [40], investigated playback quality across different device types [41] or considered scalable H.265 streaming over HTTP [42] using the Dynamic Adaptive Streaming over HTTP (DASH) [43] protocol rather than the RTP based streaming approach investigated in this paper. None of these recent H.265 QoE studies has proposed a QoE prediction model specifically aimed at the scalable extension to H.265 or the 5G context. Also, they only consider packet loss [44][38][39][45] or address a multi-user scenario where a mapping of an objective metric [46] is used to estimate QoE rather than the cost/benefit trade-off in quality offered by scalable H.265 when congestion is encountered.

III. PROPOSED 5G-QoE SYSTEM OVERVIEW

The SELFNET 5G Video QoE system provides a set of virtualised network agents within the multi-tenant SELFNET Mobile Edge Network architecture. These components, acquire, in real time, data on all video flows and the network resources through which they pass as they traverse the network from end to end. The data is aggregated and used to provide a fast, scalable and accurate estimate of the perceived quality of the video carried under the prevailing network conditions. An overview of the system architecture is shown in Fig.2.

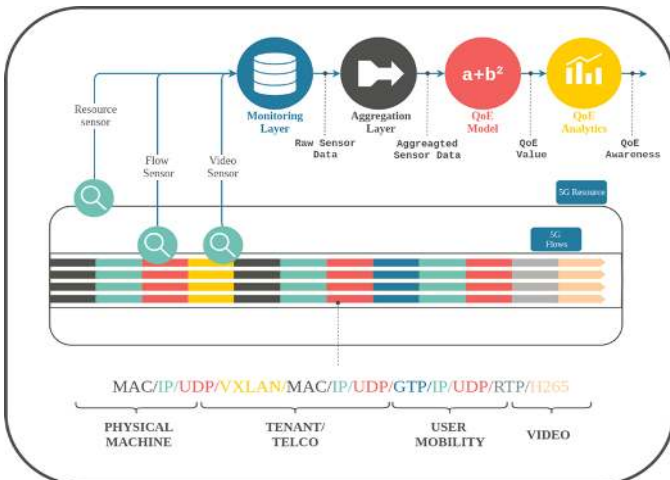


Fig.2. Overview of 5G-QoE system architecture.

The health of the 5G video transmission ecosystem is constantly monitored and analysed by periodically calculating the Quality Index of each video flow. The analysis module provides both instantaneous and time-varying QoE statistics to enable reporting or the raising of alerts on each individual video stream or set of video streams sorted by (for example) network resources such as a physical or virtual appliance or by network location (either logical or physical) or by tenant since the SELFNET platform on which this QoE system resides provides a multi-tenant infrastructure.

The analysis module can report when a video stream (or set of video streams) has fallen below the acceptable QoE threshold, or is predicted to fall below that threshold in the next reporting period, given the trend over the most recent reporting periods. It is also aware of the current adaptation state of each scalable video stream and can include in its reports information on whether a video stream has the ability to be further adapted and what the likely cost/benefit will be in terms of any trade-off between bandwidth saving and potential reduction of the Quality Index that may result from dropping one or more layers from the scalable video stream.

A. SELFNET 5G Network Monitoring Layer

The SELFNET 5G network monitoring layer consists of a three main sensors which, inspect and extract metrics from each of the components of the 5G mobile edge network (both physical and virtual) and the data flows traversing the network. Monitoring data is stored in a database, used to inform decisions in the SELFNET autonomic network management system and is also available, in aggregated form (e.g. by tenant, physical or logical zone), in SELFNET’s network management dashboard. Since the aggregated data is not associated with any subscriber data (e.g. from the home subscriber server (HSS) database), there is no impact on user privacy.

1) Flow Sensor

The SELFNET flow sensor is a virtualised 5G network agent, which inspects every flow passing through the network, acquiring information and metrics from each level of the complex set of encapsulations found in future 5G/mobile edge networks (see diagrammatic representation in Fig. 2). This sensor uniquely identifies and provides a wide range of information and metrics such as flow state (active, retired etc.), source, destination, packet count and bandwidth consumed at each level of encapsulation and tunnelling of a flow. Which in turn enables the SELFNET 5G QoE system to uniquely identify, and acquire metrics for, each layer in a scalable H.265 encoded video stream. The data is collected on a configurable periodic basis and pushed to a monitoring database. The flow sensor is a flow monitoring tool based on IPFIX protocol [56] [57]. It is hooked into the data plane by using the fast AF_PACKET [57] sniffing capabilities of the Linux kernel. This is similar to the techniques used by other well-known packet classifiers such as Wireshark [58]. AF_PACKET allows to scale the acquisition of raw packets up to 7-10 Gb/s depending on the packet size. The only difference between traditional IPfix sensors is that our flow sensor contains a customized packet classification to allow the

processing up to the double encapsulation imposed by the 5G virtualized network. Notice that IPFIX sensor is not performing any type of Deep Packet Inspection (DPI) and allows the classifier for being prototyped using an $O(n)$ linear parsing algorithm. This linear classification is analogous to the one already existing IPFIX flow sensors such as nProbe [59], Telesoft [60]. In terms of overhead, there is not any increase in the packet dropped (0%) and delay (0.000001%) over the interface where the flow sensor is running with respect to the same interfaces and traffic when the flow sensor is not running. This is due to the fact that linux kernel performs a copy of the packets into the kernel space in order to do not interference in the performance of the data plane. The overhead of the flow sensor cannot be compared against the performance of an IPFIX flow monitoring sensor since it does not provide the capabilities of the flow sensor and would be unfair to perform such comparison since traditional IPFIX sensors do not fit in virtualized 5G networks. IPFIX has recently been proposed by other authors in a similar context [61].

2) Video Sensor

Whenever a new RTP based video flow is detected by the flow sensor, it is immediately mirrored to the video sensor. Metadata, previously inserted into the stream as new Supplemental Enhancement Messages (SEI messages) by SELFNET's modified SHVC encoder (described in Section 4.2), is extracted for each scalable layer of the SHVC stream. We assumed that, these new SEI message will not be encrypted and that only the payload of packets containing other video coding layer (VCL) data is encrypted. Encoder parameters, maximum and average bitrates (for variable bitrate streams) and scene change information is gathered, associated with the unique flow identifiers for each scalable layer of the video stream and stored to the monitoring database.

3) Resource Sensor

The SELFNET 5G resource sensor monitors and acquires metrics from both the physical and virtualised infrastructures of the SELFNET 5G mobile edge network. In the context of the QoE system, metrics include configured bandwidth, current throughput and identity of data flows passing through all network interfaces within the 5G mobile edge infrastructure.

B. Data Aggregator

The data aggregation layer interrogates the monitoring database to provide aggregated performance metrics for the 5G mobile edge network. With respect to the QoE system, the first 'health of network' metric provided by the aggregator is the video flow Congestion Index (CI). CI is measures the maximum level of congestion, across all of the network interfaces a flow traverses (foreach uniquely identified layer of a scalable video stream. Both the composite SHVC video stream level CI and the CI of each individual scalable layer is calculated using the method explained in section 5. Outputs from the aggregator, (triggered by detection of a new video flow and on a configurable periodic basis) provide the initial warning of a potential reduction in QoE of users.

C. QoE Modelling

We aim to providesimple, robust and scalable QoE estimation model. The video QoE modelling agent in SELFNET provides an event driven or periodic prediction of the perceptual quality of all RTP based video streams traversing the network by combining the initial video congestion metric for each video stream with other metrics in the monitoring database. It uses inputs from the flow sensor to determine the current state of each layer in the scalable flow (e.g. active or retired), from the video sensor on layer encoding parameters such as required bitrate, spatial resolution and scene change information and available bandwidth information from the resource sensor to estimate the perceptual quality of the video stream. The prediction, known as the video Quality Index, directly maps to the Mean Opinion Score (MOS) provided by subjective evaluation experiments. The video Quality Index is SELFNET's primary 'health of network' metric for RTP based video streaming services. The amount of data required by the model is small, the QoE calculation requires the data contained in the new SEI message (typically <100 bytes), flow identifiers and the aggregated bitrates at the various interfaces that the flow traverses.

D. QoE Analysis

QoE analysis firstly establishes user tolerance of video impairments found under specific network conditions, the 5G-QoE can then analyse the 'health' of video streams crossing the network and provide appropriate QoE alerts that can be used to trigger interventions such as dropping of one or more layers from a scalable video stream to reduce network load while minimising the impact on the user's QoE.

IV. METHODOLOGY & SUBJECTIVE TESTING

A. Methodology

This section describes the methodology employed to firstly determine and subsequently validate the proposed QoE system. Firstly a set of 4k resolution video clips, with varying content types, were obtained and encoded in a scalable H.265 format, these video clips were then used in an extensive series of subjective evaluations, with a large sample size of 64, during which subjects viewed and compared both reference videos and live streamed videos where a network impairment (bandwidth limitation) had been introduced. The videos (and subjective tests) were split into two sets a training set and a validation set.

The results of the first set of subjective evaluations (training set) were used in a statistical modelling approach to derive a candidate QoE prediction formula. This formula was initially analytically validated against the subjective scores for the validation set and then subsequently implemented and empirically evaluated in the SELFNET 5G mobile edge networks testbed where all of the QoE system components described in Section 3 were used to provide empirical evidence of the effectiveness of the QoE modelling system. Fig. 3 provides a diagrammatic representation of the methodology and workflow.

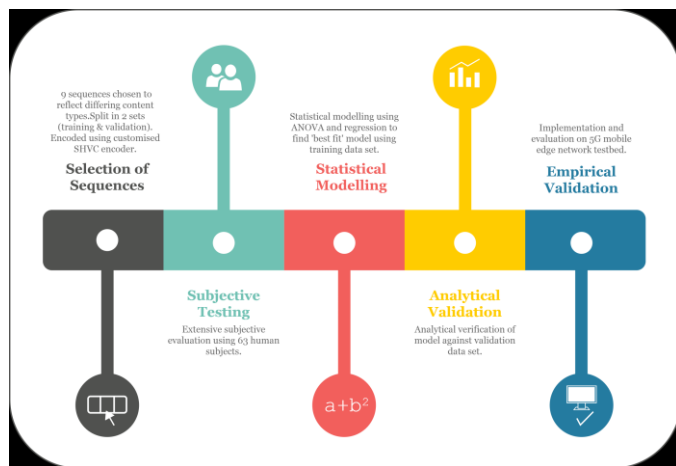


Fig.3. Diagrammatic representation of methodology and workflow.

1) Terminology used when describing experiments

To assist the reader, the terminology used in the following sections is briefly described here. Full explanations of metrics and terms are provided in the relevant sections.

- *Measured CI*: The Congestion Index reported by the 5G-QoE platform during empirical evaluations.
- *Expected CI*: The theoretical Congestion Index value expected for a given video stream/bandwidth combination. This does not take account of practical implementation issues such as network overheads or differences in reporting windows of the various 5G-QoE platform sensors in this prototype.
- *Measured QoE*: The quality of experience index calculated and reported by the 5G-QoE platform during empirical evaluations.
- *Predicted QoE*: The theoretical quality of experience index (Quality Index or Q) for a video stream/bandwidth combination provided by the QoE model.
- *Actual MOS*: The actual mean opinion score provided by human subjects through subjective testing.

B. Sequence Selection and Preparation

Firstly, nine short video sequences were encoded using a modified version of the scalable HEVC reference software (version SHM 6.1 [4]) which inserts additional NAL units containing supplemental enhancement (SEI) messages into the encoded SHVC bitstream. These messages follow guidelines for custom SEI messages in the H.265 standard and will be ignored by decoders and other network entities that do not know how to process them. The new custom SEI messages carry additional metadata to describe the maximum and average bitrate of the stream, the spatio-temporal characteristics of the stream and the frame number where a scene change occurs.

The encoder configuration employed random access

encoding and spatial scalability with two scalable layers. The standard configuration files for these encoder types were used, apart from as described below. The base layer had a spatial resolution of 1920x1080 (FHD) and a single enhancement layer with a spatial resolution of 3840x2160 (4K UHD). The clips were organised into two sets, the first set consisted of 4 clips, all with a frame rate of 30 frames per second, obtained from Ultra Video Group [48] and a second set of 5 sequences each with a frame rate of 24 fps obtained from Mitch Martinez [49]. Fig. 4 shows the spatio-temporal characteristics of the employed sequences in terms of the well-known Spatial Index (SI) and Temporal Index (TI). Fig. 4 highlights the heterogeneous nature of the sequences chosen for this study.

The sequences (including maximum and average bitrates when both scalable layers are present) are described in Table 1 and Table 2 for the 30 fps and 24 fps sequences respectively. A comparison of bitrates for each of the 9 test sequences is provided in Fig. 5, the bitrates shown are those of the SHVC encoded bitstreams (the H.265 standard compliant annex B bitstream) prior to the addition of any packet overheads required for encapsulation and transmission.

The encoded scalable bitstreams were encapsulated (in MP4 file format) and prepared for playback/transmission using version 6.1 of the GPAC framework [52], which had been compiled with the OpenHEVC decoder [53].

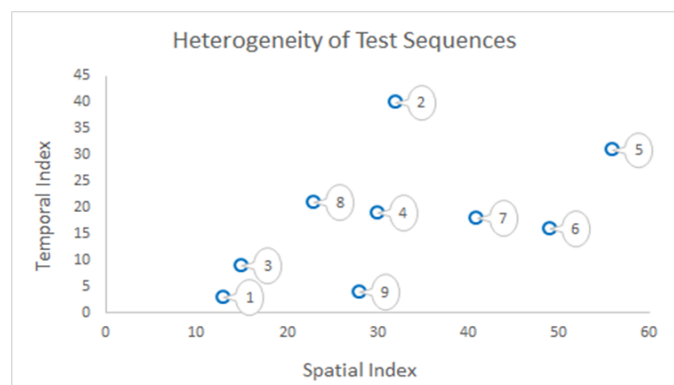


Fig. 4. Heterogeneity of test sequences.

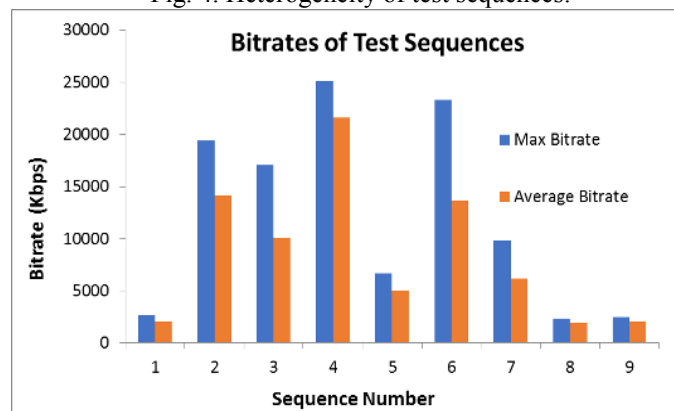


Fig. 5. Comparison of maximum and average bitrates for each sequence when both scalable layers are present.

C. Subjective Testing Platform

A bespoke testing platform (Fig. 6) facilitates subjective testing experiments where subjects compare a video sample

streamed, in real time, over an impaired network connection with a reference sample of the same video clip played locally at the client device. The platform consists of three nodes (computers) including a server node, an intermediate routing node and a client node connected to a 55-inch Samsung 4K resolution UHD television. At the server side, a console-based application manages the testing process with identical copies of the encoded video test sequences placed on both the server and the client device. The application, driven by a configuration file, synchronised the presentation of video clips to the viewers through a messaging protocol (shown in Fig. 6).

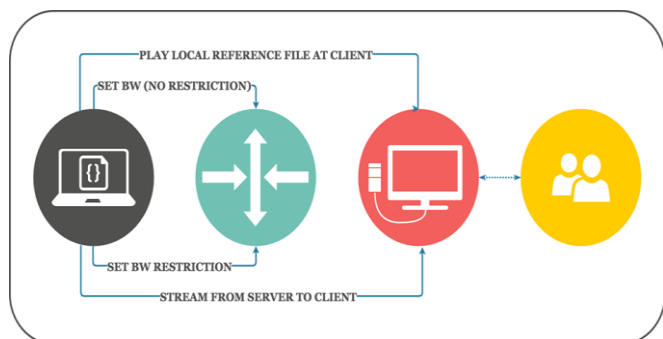
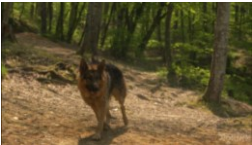





Fig. 6. Overview of the subjective testing platform.

TABLE 1.
30 FPS TEST SEQUENCES




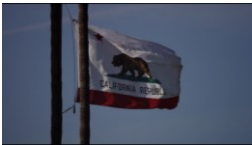

| Seq. # | Max Bitrate (Average Bitrate) (kbps) | Used For | Description & Thumbnail |
|--------|--------------------------------------|------------|---|
| 1 | 2710 (2089) | Modelling |  |
| 2 | 19445 (14125) | Modelling |  |
| 3 | 17086 (10079) | Validation |  |
| 4 | 25087 (21604) | Modelling |  |

D. Subjective Evaluation

Subjective evaluations took place over a period of two weeks in the summer of 2017. At each testing session viewers sat, in small groups of between five and seven, at the distance recommended in ITU Recommendation BT-500 [20] for the screen size. Each session lasted approximately 45 minutes with subjects able to take a break if required.

Subjects were asked to view a series of video clips, each of no more than 10 seconds duration. The clips were presented in pairs, the first of which was the reference video played locally at the client using a Scalable H.265 enabled media player [52] after a 5 second pause (during which a grey screen – as per [20] was shown), the user was then shown another copy of the same video streamed in real time from the server device to the client device and played back using the same media player.

TABLE 2.
24 FPS TEST SEQUENCES

| Seq. # | Max Bitrate (Average Bitrate) (kbps) | Used For | Description & Thumbnail |
|--------|--------------------------------------|------------|---|
| 5 | 6670 (5049) | Modelling |  |
| 6 | 23318 (13669) | Modelling |  |
| 7 | 9833 (6187) | Validation |  |
| 8 | 2334 (1925) | Validation |  |
| 9 | 2508 (2070) | Validation |  |

All video streams were transmitted using the GPAC framework from server to client over the RTP protocol. Each video clip was presented to the viewers with bandwidth limited to the maximum required by the bitstream, the average required by the bitstream, 95% of the average required and 90% of the average required.

Users were then asked to record their subjective opinion of the perceptual quality of the second video with the respect to the first using the absolute category rating scale shown in Table 3. Viewers had the opportunity to repeat any individual comparison before providing their opinion.

Initially the bandwidth between server and client devices is unrestricted (and completely sufficient to meet the requirements of the encoded bitstreams). The local reference file is played at the client, an appropriate pause (with grey screen) is given, the bandwidth on the link between client and server is applied at the egress interface of the streamer and then the same video is streamed from to the client and played back in real time. The human subjects record their opinions of the difference in quality between original and degraded videos after which the person conducting the experiments advances the platform to the next sequence to be displayed with the bandwidth restriction on the link firstly removed and then the cycle repeated for the next test sequence.

TABLE 3. ACR SCALE

| Opinion Score | Meaning |
|---------------|---|
| 5 | No difference |
| 4 | Difference perceptible but not annoying |
| 3 | Slightly Annoying |
| 2 | Annoying |
| 1 | Very Annoying |

The Mean Opinion Score derived from subjective evaluations was defined as the arithmetic mean of the scores provided by individual human subjects, as expressed in (1), where R is the rating provided by an individual user and N is the number of users:

$$MOS = \frac{\sum_{n=0}^N R_n}{N}. \quad (1)$$

E. Subjects

A total of 64 human subjects took part in the subjective evaluations, they were drawn from both University staff (lecturers and research staff) and the student body (from applied computing disciplines). 64 is a relatively large sample size in comparison to the majority of subjective evaluations presented in literature which tend to have subject numbers in the high teens or low twenties reflecting the minimum statistical values required to prove their hypothesis.

This study provides a much larger sample of both human subjects and test sequences than many other studies (which typically use four or five test sequences) aiming to provide a predictive QoE model.

The breakdown of age, sex, student to staff ratio and number of those wearing glasses is shown in Fig. 7. Of those who took part, three were excluded from the evaluation. Two were excluded as significant outliers and the third was excluded as his/her test was only partially completed. Due to the large sample sizes of human subjects (64), video test

sequences (9) and bitrate testing points (4) over 2300 individual data points were collected.

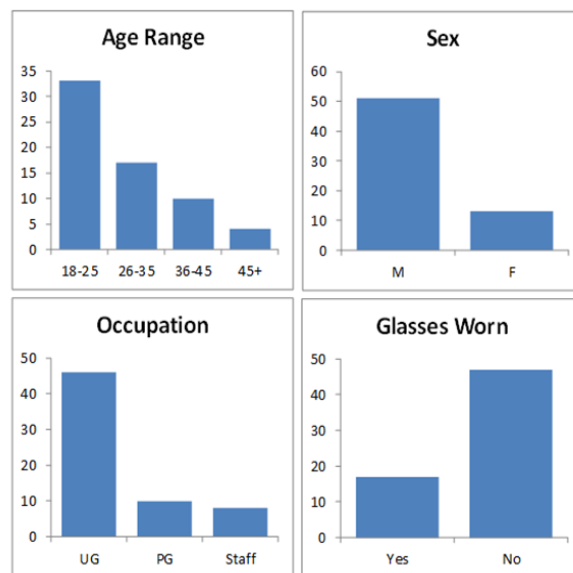


Fig.7. Breakdown of human subjects by age, sex, occupation (university staff, postgraduate students (PG) and undergraduate students (UG)) and corrected eyesight.

V. QOE MODELLING & PREDICTION

This section describes the way in which the QoE model was derived and validated. Firstly, the derivation of the video Congestion Index introduced in Section 3.2 is explained and then the statistical methods used to derive the candidate QoE index formula are discussed. After which the two validation regimes are presented together with analytical and empirical results.

The first step taken was to identify a simple network metric, with low computational overhead, which would represent the current state of the network path between sender and receiver. The Congestion Index (CI), shown in (2), is a measure of the ability of the network to successfully deliver a real time video stream based on the minimum available bandwidth on the path from sender to receiver. The Congestion Index is calculated as the ratio of the maximum required bandwidth for the stream divided by the available bandwidth for the stream:

$$CI = \frac{M}{A}. \quad (2)$$

In the subjective testing platform, the maximum bandwidth for a stream was taken to be the maximum bandwidth requirement reported by the encoder for variable bitrate videos and the available bandwidth taken to be the bandwidth limitation set on the streamer egress interface (see Fig. 6). It is worth noting

Statistical analysis of the data using one way ANOVA, regression and curve fitting produced the following function

as shown in (3) providing the best fit within the limits of the range of CI examined:

$$Q = -0.891 + \left(\frac{5.082}{\sqrt{M/A}} \right) \tag{3}$$

The subjective tests were conducted within a range of bandwidths between maximum required bandwidth and 90% of the average required bandwidth. This resulted in a range of scores where, at the higher end, their opinions indicate that the quality of video (based on the ACR scale) is either very good with little or no perceptual difference or, at the lower end, at the point where differences in quality start to become slightly annoying to the user. With this observation in mind, upper and lower bounds have been set for the values used to predict the Quality Index (i.e., predicted QoE).

```

00 import math
01 # Set the values of the constants
02 b = -0.891
03 a = 5.082
10 # Set the boundaries and triggers
11 upper_boundary = 0.8
12 lower_boundary = 1.8
13 quality_alert_trigger = 3.0
14 # calculate the Quality Index
21 congestion_index, flow_id = db_query_ci()
22 if congestion_index < upper_boundary:
23     quality_index = 5 # sufficient bandwidth to
    guarantee delivery
24 elif congestion_index > lower_boundary:
25     quality_index = 1 # no prospect of delivery at
    this bandwidth
26 else:
27     quality_index = b +
(a/math.sqrt(congestion_index))
30 #Limit the range of values
31 if quality_index > 5:
32     quality_index = 5
33 elif quality_index < 1:
34     quality_index = 1
40 #Raise alert when quality falls below acceptable
    levels
41 if quality_index < quality_alert_trigger:
42     raise_quality_alert(flow_id, quality_index)
    
```

Fig. 8 Pseudo-code for QoE prediction and alert.

As can be seen in the code example shown in Fig. 8, the upper boundary at a Congestion Index of 0.8, meaning that the bandwidth available for the video stream is at least 20% higher than the maximum bitrate of the stream. The lower boundary is set at 1.8; this figure was chosen since over 90% of subjective MOS score indicated a MOS of 3 or less at this

CI level indicating a need for some form of network intervention, by for example dropping a scalable layer, to maintain user satisfaction levels. Therefore, in the code used to implement the prototype, CI values above 1.8 automatically assume the quality will fall below the acceptable threshold.

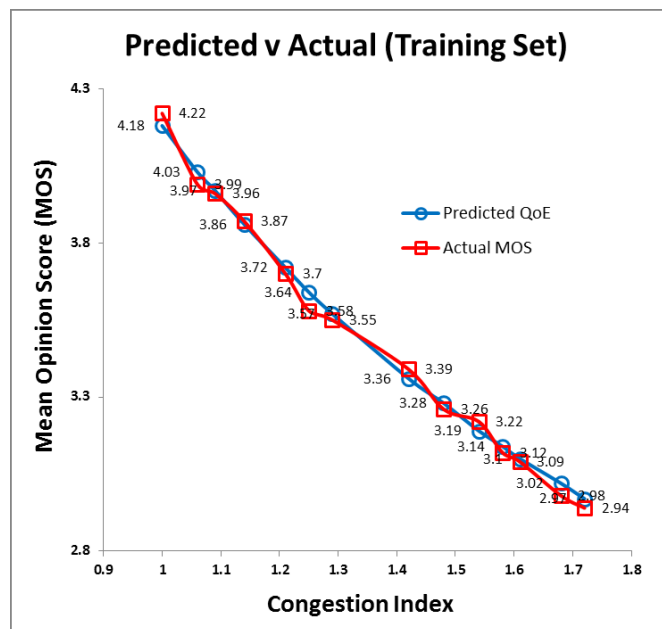


Fig. 9 Comparison of the predicted QoE with the actual MOS in the training set.

Comparisons between the predicted QoE of training set and the actual MOS perceived by the human subjects are shown in Table 4, which highlights the correlation, 0.01 significance level, and Fig. 9, which provides a graphical comparison of the same data set.

TABLE 4
CORRELATION BETWEEN THE PREDICTED QoE AND ACTUAL MOS FOR THE TRAINING SET

| | | Correlations | |
|----------------|-----------------------------------|---------------|------------|
| (Training Set) | | Predicted QoE | Actual MOS |
| Predicted QoE | Pearson Correlation | 1 | .998** |
| | Sig. (2-tailed) | | .000 |
| | Sum of Squares and Cross-products | 195.827 | 198.416 |
| | Covariance | .152 | .154 |
| | N | 1287 | 1287 |
| Actual MOS | Pearson Correlation | .998** | 1 |
| | Sig. (2-tailed) | .000 | |
| | Sum of Squares and Cross-products | 198.416 | 202.031 |
| | Covariance | .154 | .157 |
| | N | 1287 | 1287 |

** . Correlation is significant at the 0.01 level (2-tailed).

VI. QoE MODEL VALIDATION

A. Analytical Validation

Having derived the QoE prediction formula using the subjective evaluation data of the five training set video sequences, the formula was then analytically validated by comparing the predicted output of the QoE formula for each of the four validation sequences with the actual Mean Opinion Scores provided by human subjects during the subjective tests. The Congestion Index was calculated as shown in (2). The available bandwidth used in this equation was the bandwidth restriction applied at the client to server link of the subjective testing platform. The same bandwidth restriction ratios, as used in the subjective testing, were employed to ensure a fair and accurate comparison.

The results of this analytical comparison are shown in the following two subsections: one describes the results of the analytical validation when both layers of the scalable bitstream were present, and the other presents validation results for the H.265 base layer only.

When both layers were present in the bitstream the analytical validation shows a close correlation between the predicted QoE by the formula (3) and the actual MOS from subjective testing. In Fig. 10, the average MOS scores of all testing points for the validation set of video sequences are compared with the predicted QoE for each testing point.

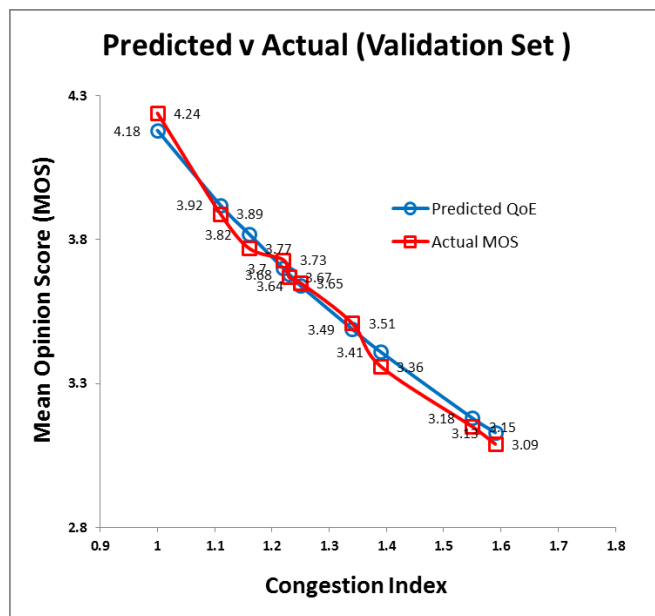


Fig. 10 Analytical validation by comparing the predicted QoE with the actual MOS in the validation set.

From Fig.10, it can clearly be seen that the predicted QoE results closely track the average subjective MOS scores provided by human subjects. Furthermore, it can also be seen from Fig. 11 that the variance between the actual subjective test MOS scores and the predicted QoE is in the range ± 0.06 for both the training set and the validation set. It can also be seen from Table 5 that the correlation between the predicted

QoE and actual MOS is very high.

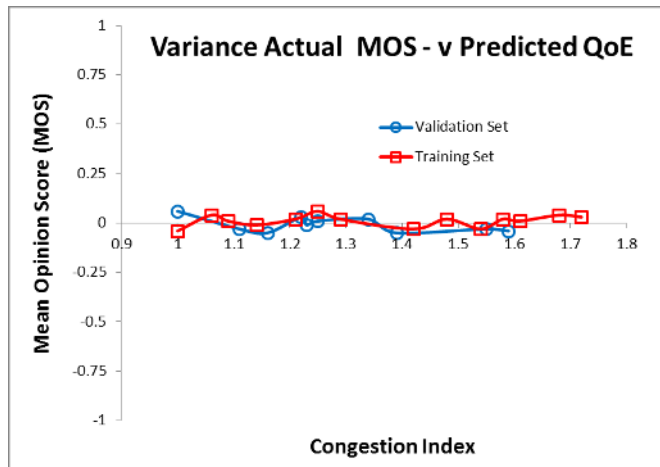


Fig. 11 Analytical validation by comparing the predicted QoE with the actual MOS in the training and validation sets – variance.

Another interesting result from the subjective testing, shown in Table 6 is that there is very little difference in perception of quality between 4K and FHD versions of a video when streamed with sufficient bandwidth. In fact, some viewers found the FHD version to be of better quality than the 4K version. This supports our assertion that dropping a scalable layer will have little impact on the user’s perception of quality.

TABLE 5
CORRELATION BETWEEN PREDICTED QoE AND ACTUAL MOS FOR THE VALIDATION SET

| | | Correlations | |
|------------------|-----------------------------------|---------------|------------|
| (Validation Set) | | Predicted QoE | Actual MOS |
| Predicted QoE | Pearson | 1 | .996** |
| | Correlation Sig. (2-tailed) | | .000 |
| | Sum of Squares and Cross-products | 104.330 | 110.583 |
| | Covariance | .095 | .101 |
| | N | 1100 | 1100 |
| | Actual MOS | Pearson | .996** |
| Actual MOS | Correlation Sig. (2-tailed) | .000 | |
| | Sum of Squares and Cross-products | 110.583 | 118.232 |
| | Covariance | .101 | .108 |
| | N | 1100 | 1100 |

** . Correlation is significant at the 0.01 level (2-tailed).

TABLE 6
COMPARISON OF 4K AND FHD WHEN CI=1 FOR THE VALIDATION SET

| Resolution | Mean MOS | StdDev |
|------------|----------|--------|
| 4K | 4.24 | 0.23 |
| FHD | 4.21 | 0.26 |

B. Empirical Validation

After analytically establishing that the proposed QoE prediction model provided a close approximation of the actual Mean Opinion Scores of human subjects, the QoE model was further evaluated and validated by implementing it on an experimental 5G-QoE testbed. This took the form of a two-stage process where video test sequences were firstly prepared to generate 5G video traffic, before being evaluated on a dedicated 5G-QoE testing platform.

1) Preparing sequences for empirical evaluation

As part of the SELFNET project, a 5G Infrastructure Testbed has been established to represent a realistic end-to-end 5G mobile network infrastructure comprising a Radio Access Network (RAN) and a core network. This 5G Infrastructure Testbed, together with the positioning of the 5G-QoE system, is illustrated in Fig. 12.

This testbed was implemented using the open source OpenAirInterface implementation [50], and the 5G infrastructure has been achieved through introducing the Cloud-RAN model and core network virtualization following the LTE (Long Term Evolution) evolution based 5G realization approach. In the RAN and core networks, the key LTE components have been virtualized accordingly, including BBU (BaseBand Unit), MME (Mobility Management Entity), HSS (Home Subscriber Server), PGW (Packet Data Network Gateway), and SGW (Serving Gateway). PCRF (Policy and Charging Rules Function) is not implemented yet as it does not affect the studies; however, it is illustrated in Fig. 12 for completeness. Further details on this platform can be found at [55]. This testbed enables studies of the 5G infrastructure side but has a low capacity of the air interface, not suitable for rigorous UHD video testing. To circumvent this limitation, a dedicated 5G-QoE Testbed has been created in this research to compensate for the low capacity of the air interface in the first testbed’s RAN.

To this end, in this preparation stage, the video test sequences were firstly streamed across the 5G Infrastructure Testbed to obtain realistic 5G network traces (in the form of PCAP files) containing all of the encapsulation layers found in a multi-tenant 5G mobile network infrastructure. The encapsulation layers are illustrated in Fig. 2. This approach facilitates the later on emulation of the full end-to-end mobile infrastructure including 5G air interface on the standalone 5G-QoE Testbed.

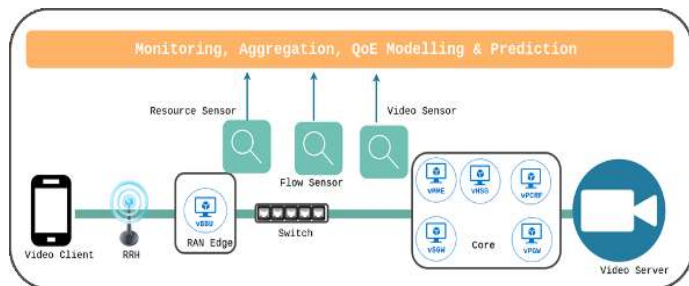


Fig. 12. 5G Infrastructure Testbed with the positioning of

the 5G-QoE system to create and capture realistic 5G traffic.

2) Empirical 5G-QoE Testbed

Fig. 13 illustrates the overview of the 5G-QoE Testbed deployed to conduct empirical validation of the proposed system. Two physical machines are used including one laptop to act as the UHD video streamer, and one high-end PC to host the rest of the system (the various software sensors, and the monitoring, aggregation, QoE modelling and prediction and analysis software modules), as described in Section 3.

The Streamer machine is installed with Ubuntu 16.04 LTS 64-bit operating system, and equipped with 8 GiB RAM, a 10 core (4 x compute, 6 x graphics) AMD A10-7300 Radeon R6 processor, and a 1.0 TB hard disk. The PC runs Ubuntu 14.04 LTS 64-bit OS and features 32 GiB RAM, a 16 core Intel Xeon(R) CPU-E5-2630 v4 @ 2.20GHz processor, and a 2.0 TB hard disk.

The testbed is configured in such a way that it can either stream and playback 4K UHD video encoded with the Scalable HEVC (or standard HEVC) encoder over a standard IPv4 network connection or it can use tcpreplay [54] to transmit the previously captured 5G PCAP files from the streamer to generate realistic 5G video flows.

The ten testing points shown in Fig 15 and Fig 16 were averages values obtained by using validation sequences in the live testbed. A range of CI values from 1.0 to 1.60 (1, 1.11, 1.16, 1.22, 1.23, 1.25, 1.34, 1.39, 1.55, 1.59,) corresponding to the letters A to J.

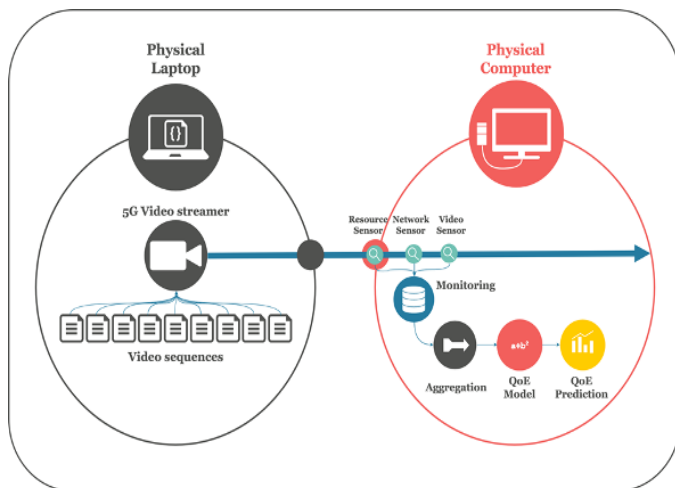


Fig. 13 5G-QoE testbed for empirical experiments.

3) Empirical Validation Results

The purpose of the empirical validation was firstly to demonstrate that the QoE prediction model could be successfully implemented as part of a prototype 5G QoE system, in which all of the components described in section 3 work in unison. Secondly, to deliver a system that is sufficiently lightweight to inform real time network management decisions, yet accurate enough to provide a realistic approximation of the QoE level that a user can expect

under the prevailing network conditions.

Fig. 14 provides an example of the output from a periodic health of video network report from the analysis component of the 5G-QoE system shows the estimates of both Congestion Index (CI) and Quality Index (MOS) for an H.265 scalable video stream containing 2 scalable layers.

The upper line of the output estimates the CI and Quality Index when both layers are present and the lower line those if only the base layer is present. It can be observed that, in this example, the Quality Index when both layers are present suggests that users would typically find the video quality slightly annoying, whilst if only the base layer were present they would find the quality to be acceptable (perceptible but not annoying). It is also worth noting that, in practical terms, the variability in user perceptions of quality mean that an average MOS score of 4.3 from our subjective testing indicates a high level of satisfaction with the quality of video.

| nicSpeed | currentBandwidthConsumed | flowVideoCongestionIndex | flowVideoMOS |
|----------|--------------------------|--------------------------|--------------------|
| 2500000 | 1249259.1111111111 | 1.4434560965059318 | 3.3389269879418615 |
| 2500000 | 1249259.1111111111 | 0.9465600832105564 | 4.332487594715371 |

Fig. 14 Sample of measured consumed bandwidth, CI and QoE values at different layers in the empirical tests.

Given that the proposed QoE prediction model is in fact a relatively simple function of the video flow Congestion Index (CI) presented in (2), the first step was to validate the accuracy of the empirically reported CI from the testbed against the expected CI, which had been analytically derived from the maximum and average bitrates required by the encoded bitstreams (as reported by the scalable H.265 video encoder SHVC). This initial validation step, which may in itself appear to be no more than a trivial comparison, in fact validated the accuracy and effectiveness of each of the complex set of interacting components of the 5G-QoE system. These components (network sensor, video sensor, data aggregator, messaging bus used for data collection and the database acquired metrics and flow statistics) all need to work in unison to provide the data input (CI) to the QoE prediction model.

Fig. 15 provides a graphical comparison between the expected CI (analytically derived) and the measured CI obtained during empirical validation. Each of the testing points (indicated by the test number) was created by varying the available bandwidth and video sequences (having various maximum required bandwidth conditions), and was evaluated three times with the mean CI for each testing point reported in the graph. It can be seen that the empirically obtained CI, whilst not identical, closely tracks the expected CI. These results show that the monitoring and aggregation components of the 5G-QoE system performed as expected.

The modest difference in CI can be explained by several factors. Firstly, the sensors and aggregator each have tuneable reporting frequency parameters (set to 1 second during this evaluation) which, when combined with the variable nature of

the encoded bitstream, could lead to some discrepancy in the reporting of the currently consumed bandwidth of the video stream. Secondly, the subjective evaluation platform and analytical evaluation were based on the pure IPv4 transmissions used in the subjective testing platform, whereas in the empirical evaluation the full 5G/LTE PCAP files containing all of the encapsulation layers were used. Although this was compensated for by counting the average number of packets per reporting period and increasing the maximum required bitrate of the stream used in the CI calculation to take account of all encapsulation overheads, again the variable nature of the bitstream meant that some small margin of error could be expected in this calculation.

Taken as an average across all of the empirical evaluations, the empirically observed CI varied from the expected CI within the range of $\pm 6\%$, and we consider this an acceptable margin of error for a prototype implementation.

Turning to the comparison between actual (subjective) MOS, predicted QoE and empirically measured QoE, the results again show a strong correlation between each set of results. A comparison of MOS/QoE scores is provided in Fig. 16, which shows that scores for empirically measured QoE closely track those for both predicted QoE (from the QoE prediction model) and the actual MOS scores from subjective testing. The X-axis values in this figure align with those in Fig. 15.

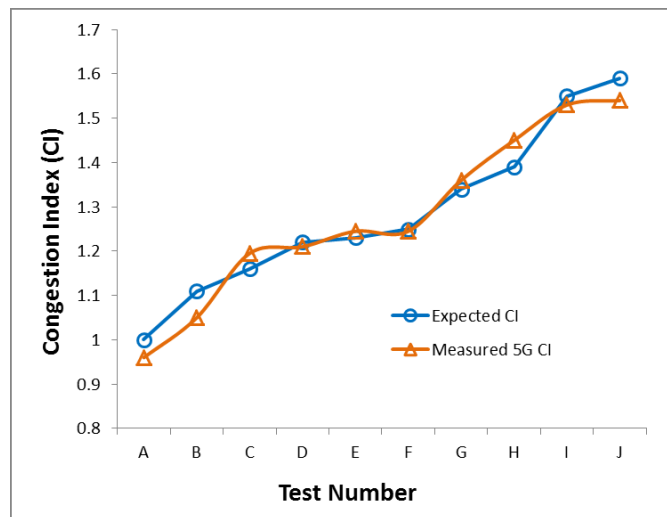


Fig. 15 Empirical validation by comparing expected CI with measured CI (vs. test number).

Fig. 17, which allows easy comparison with Fig. 10, provides a plot of mean opinion score (Y-axis) against Congestion Index (X-axis) for predicted QoE and empirically measured QoE. Again the differences shown are modest, with the largest variation between predicted and measured QoE recorded as a 0.06 difference in average scores (of the three tests at each testing point).

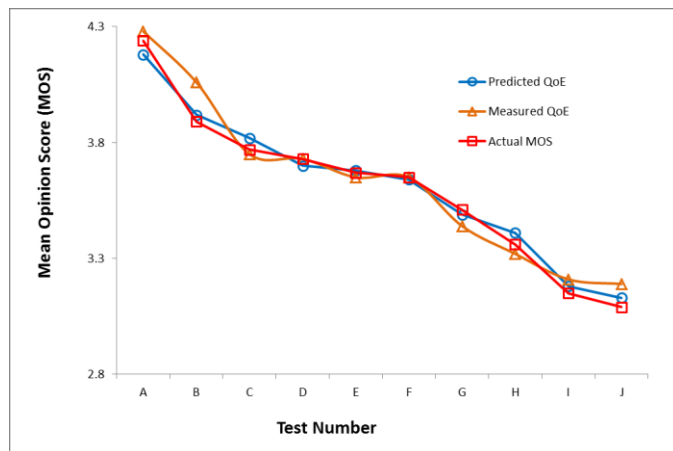


Fig. 16 Empirical validation by comparing predicted QoE with measured QoE and actual MOS (vs. test number).

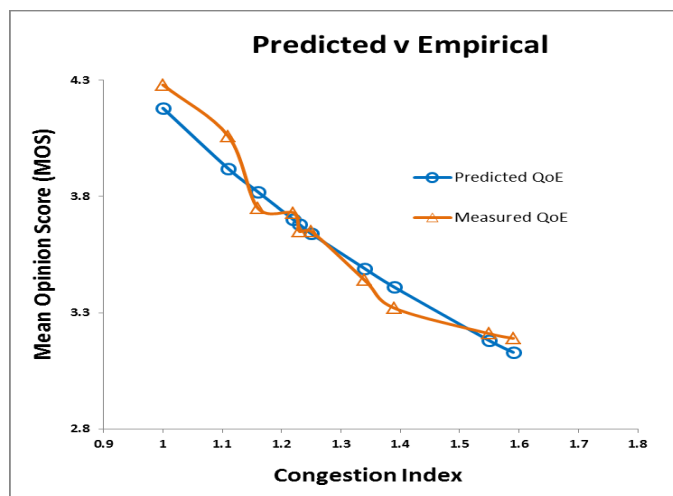


Fig. 17 Empirical validation by comparing predicted QoE with measured QoE (vs. CI).

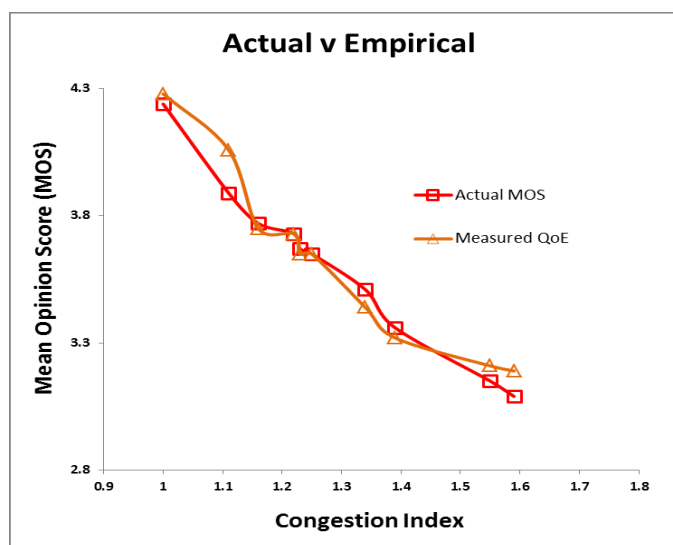


Fig. 18 Empirical validation by comparing measured QoE with actual MOS (vs. CI).

VII. CONCLUSION

This paper has presented a fast and scalable method of estimating the perceived quality of experience of users of UHD video flows in the emerging 5G networks as part of a comprehensive 5G-QoE framework. The model has been analytically and empirically evaluated against the results of subjective testing with results showing an accuracy of up to 94%. The 5G-QoE framework has been implemented on the EU 5G PPP SELFNET platform, where the model has been demonstrated to work as part of the SELFNET mobile edge infrastructure, taking account of all tunnelling overheads introduced to the video flows by 5G infrastructure to achieve multi-tenancy and mobility, and providing empirical QoE scores that closely match both those predicted by the model and actual MOS scores of the test subject, with the maximum variance only 0.06 and 0.17 respectively.

Future work will concentrate on building a QoE-aware video adaptation system that leverages the 5G-QoE framework to analyse and optimize likely user perception of quality for scalable H.265 encoded UHD video streams. This system will act as a first line of defence and will inform decisions for smart traffic engineering, for example, when and which layers of a scalable video stream should be dropped in the concerned network congestion situations in order to maximise benefit to network operations while minimising the impact on perceived QoE.

ACKNOWLEDGMENT

This work was funded in part by the European Commission Horizon 2020 5G PPP Programme under grant agreement number H2020-ICT-2014-2/671672 – SELFNET (Self-Organized Network Management in Virtualized and Software Defined Networks). The authors wish to thank all the SELFNET partners for their support in this work. This work was additionally funded by the UWS 5G Video Lab project. The authors would also like to thank Mitch Martinez for giving permission to use his library of 4K video clips.

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update", White Paper, 2016, [online] Available: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visualnetworking-index-vni/mobile-white-paper-c11-520862.html>.USA.
- [2] 5G PPP, "Advanced 5G Network Infrastructure for the Future Internet," 2013. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2014/02/Advanced-5GNetwork-Infrastructure-PPP-in-H2020_Final_November-2013.pdf.
- [3] 5G PPP, "5G Vision - The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services", [Online]. Available at <https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf>
- [4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "https://doi.org/10.1109/TCSVT.2012.2221192", IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
- [5] ITU-T, Recommendation H.265 (V4), High efficiency video coding, Dec. 2016, [online] Available: <http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=11885>.

- [6] DisplayPort, available online, http://www.displayport.org/wp-content/uploads/2016/06/Display-Summit-USA-2016-AMD-VESA-SAH_Final_rl.pdf
- [7] Japan Display, Press Release, 2017, <http://www.j-display.com/english/news/2017/20170519.html>
- [8] 5G PPP, "5G-PPP Key Performance Indicators (KPIs)," [Online]. Available at <https://5g-ppp.eu/kpis/DaunmuDaunmu>
- [9] The EU 5G PPP SELFNET project, "SELFNET: a framework for self-organized network management in virtualized and software defined networks", Apr. 2017, [online] Available: <https://selfnet-5g.eu/>.
- [10] P. Neves et al., "Future Mode of Operations for 5G - The SELFNET Approach Enabled by SDN/NFV", (Elsevier) Computer Standards & Interfaces, in press (accepted). (doi: <http://dx.doi.org/10.1016/j.csi.2016.12.008>)
- [11] Takahashi A., Hands D. and Barriac V., "Standardization activities in the ITU for a QoE assessment of IPTV", IEEE Communications Magazine, Vol. 46, No. 2, February 2008. doi.org/10.1109/MCOM.2008.4473087
- [12] I. Slivar, M. Suznjevic and L. Skorin-Kapov, "The impact of video encoding parameters and game type on QoE for cloud gaming: A case study using the steam platform," 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX), Pylos-Nestoras, 2015, pp. 1-6.
- [13] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp. 1103–1120, Sep. 2007
- [14] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramanian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 20-34, Jan. 2016.
- [15] Netflix, 2017, [online] Available: <https://www.netflix.com/>
- [16] Hulu, 2017, [online] Available: <https://www.hulu.com/>
- [17] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," IEEE Trans. Broadcast., vol. 57, no. 2, pp. 165182, Jun. 2011.
- [18] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hofbeld 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, First quarter 2015.
- [19] Z. Akhtar, T. H. Falk, "Audio-Visual Multimedia Quality Assessment: Comprehensive Survey", IEEE Access, Vol. 5Takahashi, Sept 2017, pp. 21090 – 21117
- [20] ITU, "Methodology for the subjective assessment of television pictures," ITU-R BT500-13, Jan. 2012
- [21] U. Reiter et al., "Factors influencing quality of experience," in Quality of Experience: Advanced Concepts, Applications and Methods. Cham, Switzerland: Springer, 2014, pp. 55–72
- [22] P. A. Kara, L. Bokor, A. Sackl, and M. Mourão, "What your phonemakes you see: Investigation of the effect of end-user devices on the assessment of perceived multimedia quality," in Proc. Int. Workshop Qual. Multimedia Exp. (QoMEX), Pylos, Greece, May 2015, pp. 1–6
- [23] M. Alreshoodi, E. Danish, J. Woods, A. Fernando and C. De Alwis, "Prediction of Perceptual Quality for Mobile Video Using Fuzzy Inference Systems," IEEE Transactions on Consumer Electronics 2015, vol. 61, no. 4, pp. 546–554.
- [24] M. Seydebrahimi, C. Bailey, and X.-H. Peng, "Model and Performance of a No-Reference Quality Assessment Metric for Video Streaming," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 12, Dec. 2013, pp. 2034-2043.
- [25] W. Hsu and C. Lo, "QoS/QoE Mapping and Adjustment Model in the Cloud-Based Multimedia Infrastructure," IEEE Syst. J., vol. 8, no. 1, 2014, pp. 247–55.
- [26] J. A. Khan, L. Sun, E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks", IEEE Trans. Multimed., vol. 14, no. 2, pp. 431-442, Apr. 2012.
- [27] The 5G Infrastructure Public Private Partnership (5G-PPP), "First Wave of Research & Innovation Projects", 2016, [online] Available: <https://5g-ppp.eu/wp-content/uploads/2015/10/5GPPP-brochure-final-web.pdf>.
- [28] EU 5G PPP Architecture Working Group, "5G PPP View on 5G Architecture", Jun/Jul 2016, available at <https://5g-ppp.eu/white-papers/>.
- [29] EU 5G PPP Software Networking Working Group, "Vision on Software Networks and 5G", Jan 2017.
- [30] UHD-on-5G project, 2016, [online] Available: <https://team.inria.fr/diana/uhd-on-5g/>.
- [31] M. A. Kourtis, H. Koumaras, G. Xilouris and F. Liberal, "An NFV-based Video Quality Assessment Method over 5G Small Cell Networks," IEEE MultiMedia, vol. 24, no. 4, Oct-Dec 2017, pp. 68-78.
- [32] C. Ge, N. Wang, G. Foster and M. Wilson, "Toward QoE-Assured 4K Video-on-Demand Delivery Through Mobile Edge Virtualization With Adaptive Prefetching," IEEE Transactions on Multimedia, vol. 19, no. 10, pp. 2222-2237, Oct 2017.
- [33] 3GPP, "Evaluation of High Efficiency Video Coding (HEVC) for 3GPP services (Release 12)", TR 26.906 V12.0.0, 2014.
- [34] NGMN Alliance, "Perspectives on Vertical Industries and Implications for 5G", White Paper, Jun. 2016, [online] Available: https://www.ngmn.org/uploads/media/160610_NGMN_Perspectives_on_Vertical_Industries_and_Implications_for_5G_v1_0.pdf.
- [35] <https://doi.org/10.1109/TCSVT.2012.2221192>
- [36] S. Radicke, J. Hahn, Q. Wang, and C. Grecos, "A Parallel HEVC Intra Prediction Algorithm for Heterogeneous CPU+GPU Platforms", IEEE Transactions on Broadcasting, Vol. 62, No. 1, Mar 2016, pp. 103-119
- [37] J. Nightingale, Q. Wang, C. Grecos, S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming", IEEE Trans. Consumer Electron., vol. 60, no. 2, pp. 242-250, May 2014
- [38] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "Video Adaptation for Consumer Devices: Opportunities and Challenges Offered by New Standards", IEEE Communications Magazine, Vol. 52, No. 12, Dec 2014.
- [39] Y. Ye, Y. He and X. Xiu, "Manipulating Ultra-High Definition Video Traffic," in IEEE MultiMedia, vol. 22, no. 3, pp. 73-81, July-Sept. 2015
- [40] Paudyal, P., Battisti, F. & Carli, M. Multimed Tools Appl (2016) 75: 16461. <https://doi.org/10.1007/s11042-015-3214-0>
- [41] R. Garcia and H. Kalva, "Subjective evaluation of HEVC and AVC/H.264 in mobile environments," in IEEE Transactions on Consumer Electronics, vol. 60, no. 1, pp. 116-123, February 2014
- [42] Z. Duanmu, K. Zeng, K. Ma, A. Rehman and Z. Wang, "A Quality-of-Experience Index for Streaming Video," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 1, pp. 154-166, Feb. 2017
- [43] Information Technology—Dynamic Adaptive Streaming Over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats, ISO/IEC 23009-1:2012, 2012.
- [44] L. Qian, Z. Cheng, Z. Fang, L. Ding, F. Yang and W. Huang, "A QoE-Driven Encoder Adaptation Scheme for Multi-User Video Streaming in Wireless Networks," in IEEE Transactions on Broadcasting, vol. 63, no. 1, pp. 20-31, March 2017
- [45] Ryu, ES. & Ryu, S. Multimed Tools Appl (2017) 76: 25511. <https://doi.org/10.1007/s11042-017-4835-2>
- [46] A. N. Moldovan, I. Ghergulescu and C. H. Muntean, "VQAMap: A Novel Mechanism for Mapping Objective Video Quality Metrics to Subjective MOS Scale," in IEEE Transactions on Broadcasting, vol. 62, no. 3, pp. 610-627, Sept. 2016
- [47] J. Joskowicz, R. Sotelo, J. C. L. Ardao, "Towards a general parametric model for perceptual video quality estimation", IEEE Trans. Broadcast., vol. 59, no. 4, pp. 569-579, Dec. 2013.
- [48] UltraVideo Group, 4K Test Sequences, available online, <http://ultravideo.cs.tut.fi/#testsequences>
- [49] Mitch Martinez, 4K video sequences, available online at <http://mitchmartinez.com/free-4k-red-epic-stock-footage/>
- [50] OpenAirInterface, 2017, [online] Available: <http://www.openairinterface.org/>
- [51] SHVC reference software, 2017, [Online]. Available: <https://hevc.hhi.fraunhofer.de/shvc>.
- [52] GPAC, 2017, [Online]. Available: <https://gpac.wp.imt.fr/home/>
- [53] OpenHEVC, 2017, [Online]. Available: <https://github.com/OpenHEVC/openHEVC>
- [54] Tcpreplay, 2017, [Online]. Available: <http://tcpreplay.appneta.com/>
- [55] SELFNET, Deliverable 2.4: Portable Testbed to Execute Virtualized NFV-based and SDN-based Scenarios, [online]. Available https://bscw.selfnet-5g.eu/pub/bscw.cgi/d37407-4/*/*/*/*DOI-D2.4.html
- [56] B. Claise, B. Trammell, and P. Aitken, "Specification of the IP flow information export (IPFIX) protocol for the exchange of flow information," IETF, Fremont, CA, USA, RFC 7011, Sep. 2013. [Online] Available: <http://www.ietf.org/rfc/rfc7011.txt>
- [57] B. Trammell and E. Boschi, "An introduction to IP flow information export (IPFIX)," IEEE Commun. Mag., vol. 49, no. 4, pp. 89–95, Apr. 2011. [Online]. Available: <http://dx.doi.org/10.1109/MCOM.2011.5741152>
- [58] Wireshark packet analyser [Online]. Available <https://www.wireshark.org/>

- [59] nProbe extensible network probe [Online]. Available <https://www.ntop.org/products/netflow/nprobe/>
- [60] Telsogt IP flow probe, [Online]. Available <https://www.telesoft-technologies.com/cyber/ip-flow-probe>



James Nightingale is a Postdoctoral Research Fellow with the University of the West of Scotland (UWS), UK. He is working on the EU H2020 5G-PPP Project SELFNET, and has worked on other projects such as the UK EPSRC project “Enabler for Next-Generation Mobile Video Applications”. His research interests include mobile networks and video streaming techniques. He has a PhD in mobile video streaming from UWS, UK.



Pablo Salva-Garcia is a PhD candidate at the University of the West of Scotland, UK, where he is involved in the H2020 5G-PPP Phase I SELFNET project. His main interests include network management, cognitive control plane, software data paths, software-defined networks and video delivery in mobile edge computing and 5G networks.



Jose M. Alcaraz Calero is a Full Professor in networks and security at the University of the West of Scotland, and he is the technical co-coordinator of the EU H2020 5G-PPP Phase I SELFNET and Phase II SliceNet. His professional interests include network cognition, management, security and control, service deployment, automation and orchestration, and 5G mobile networks. He has a PhD in computer Science, University of Murcia,

Spain.



Qi Wang is a Full Professor at the University of the West of Scotland, and he is the technical co-coordinator of EU H2020 5G-PPP Phase I SELFNET and Phase II SliceNet. He is a Board Member of the Technology Board of EU 5G-PPP, and Member of USA Video Quality Expert Group (VQEG). His research primarily focuses on 5G mobile networks and video networking. He has a PhD in mobile networking from the University of Plymouth, UK.