5G-QOE: QOE MODELLING FOR ULTRA-HD VIDEO STREAMING IN 5G NETWORKS

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

ANAND CHANDAR P B

KNOWLEDGE IS POWER

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering
Kothamangalam

5G-QOE: QOE MODELLING FOR ULTRA-HD VIDEO STREAMING IN 5G NETWORKS

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

ANAND CHANDAR P B

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering
Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the seminar report entitled **5G-QoE : QoE Modelling For Ultra - HD Video Streaming In 5G Networks** *submitted by* **Mr. ANAND CHANDAR P B** , *Reg.No.* **MAC15CS011** *towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for December 2018 is a bonafide record of the seminar carried out by him under our supervision and guidance.*

....	....	....
Prof. Joby George	**Prof. Neethu Subash**	**Dr. Surekha Mariam Varghese**
Faculty Guide	***Faculty Guide***	***Head Of Department***

Date: Dept. Seal

Acknowledgment

First and foremost, I sincerely thank the 'God Almighty' for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Solly George, Principal and Dr. Surekha Mariam Varghese, Head Of the Department for providing the necessary facilities and their encouragement and support.

I owe special thanks to the staff-in-charge Prof. Joby george, Prof. Neethu Subash and Prof. Joby Anu Mathew for their corrections, suggestions and sincere efforts to co-ordinate the seminar under a tight schedule.

I express my sincere thanks to staff members in the Department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally, I would like to acknowledge the heartfelt efforts, comments, criticisms, co-operation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish.

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. The paper proposes a 5G-QoE framework to address the QoE modeling for UHD video flows in 5G networks. 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. The model has been developed and implemented as part of the EU 5G PPP SELFNET autonomic management framework. 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.

Contents

List of Figures

List of Abbreviations

HEVC : High Efficiency Video Coding

UHD : Ultra High Definition

QOE : Quality of Experience

CDN : Content Delivery Networks

MOS : Mean Opinion Score

ACR : Absolute Category Rating Scale

PSNR : Peak-Signal-to-Noise Ratio

SSIM : Structural Similarity Index Matrix

CI : Congestion Index

LTE : Long Term Evolution

BBU : BaseBand Unit

MME : Mobility Management Entity

PGW : Packet Data Network Gateway

SGW : Serving Gateway

RAN : Radio Access Network

Introduction

Fifth Generation (5G) mobile networks enter service with expected higher bandwidths, lower end to end delays and improved reliability, are likely to increase demand for mobile video consumption. New video compression standards such as High Efficiency Video Coding (H.265/HEVC) [1], [2] and the availability of Ultra-High-Definition (UHD) portable consumer devices may increases the growth in mobile video traffic. 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. 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.

The EU 5G PPP SELFNET project [3] 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.

- Finding a reliable, accurate, scalable and robust QoE prediction model for streamed video over mobile networks is an unresolved and very challenging task.
- Current QoE models including those promoted by standardisation bodies [4] not focus on 5G networks where additional challenges such as virtualisation, mobility and multi-tenancy requirements exist.
- Video encoder type is a significant factor in QoE modelling , existing QoE models usually only consider single layer video encoders mostly for the H.264 Advanced Video Coding standard (H.264/AVC).

To address these challenges, an investigation is done as a QoE prediction of UHD video, encoded using the scalable extension to the H.265 standard (SHVC) [5] 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

This work addresses real-time RTP-based video streaming often used for video conferencing, video chat and video surveillance applications rather than the Dynamic Streaming over HTTP (DASH) based streaming used in Content Delivery Networks (CDNs) such as Netflix or Hulu 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

The remainder of this work is organised as follows. The next section reviews the state of the art in QoE modelling for streamed video, scalable video codecs where advances towards autonomic functionality in 5G networks. Following that, provides an insight into the QoE-driven, self-optimising features of the SELFNET 5G network management architecture.Then explains the methodology used and the subjective testing experiments undertaken. After that a QoE prediction model is developed and the results of validation experiments presented. The Final section draws a conclusion.

Existing Work

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.

2.1 Subjective QoE Model

QoE measurement can be performed using subjective and objective tests. Subjective tests involve direct data collection from users. For example, in the form of user ratings. Standardization bodies such as the ITU-T in its ITU-T P.800 recommendation [42] presents a methodology for conducting subjective tests. This recommendation also denes a method to measure users' QoE based on a score called the mean opinion score (MOS). MOS is used widely for subjective voice/video quality assessment where human test subjects grade their overall experience on the Absolute Category Rating Scale (ACR).

There are several problems that arise while conducting subjective tests. These tests can be expensive and time consuming. Hence, subjec- tive tests are mainly limited to major telecommunication providers. The biggest problem with MOS is that an average of user ratings is computed. Mathematical operations such as computing mean and standard deviation cannot be applied on subjective ratings as these ratings are categorical in nature (e.g., "excellent" and "fair"). The human test subjects ranks the alternatives on the categorical scale where the distance between these alter- natives cannot be known. Hence, mathematical operations cannot be applied.

2.2 Objective QoE Model

Due to 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.

Objective QoE models derive a predicted MOS from a model that is a function of some number of objectively measured parameters.And map this measurements to obtain a QoE prediction model. evaluation. 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. These parameters have often included 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 QoE metrics.

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 further empirically validated in a realistic 5G testbed.

Proposed Work

SELFNET 5G-QOE SYSTEM

The SELFNET 5G-QoE system provides a set of virtualised network agents. These components acquire real time data on video flows and the network resources through which they pass as they traverse the network from end to end. The data are aggregated and used to provide a fast, scalable and accurate estimation of the perceived quality of the video carried under the prevailing network conditions. An overview of the system architecture is shown in Fig. 3.1. The health of the 5G video transmission is constantly monitored and analysed by periodically calculating a 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 network resources such as a physical or virtual machine or by network location (either logical or physical).

The analysis module can report when a video stream has fallen below the acceptable QoE threshold, or is predicted to fall below that threshold in the next reporting period. 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.

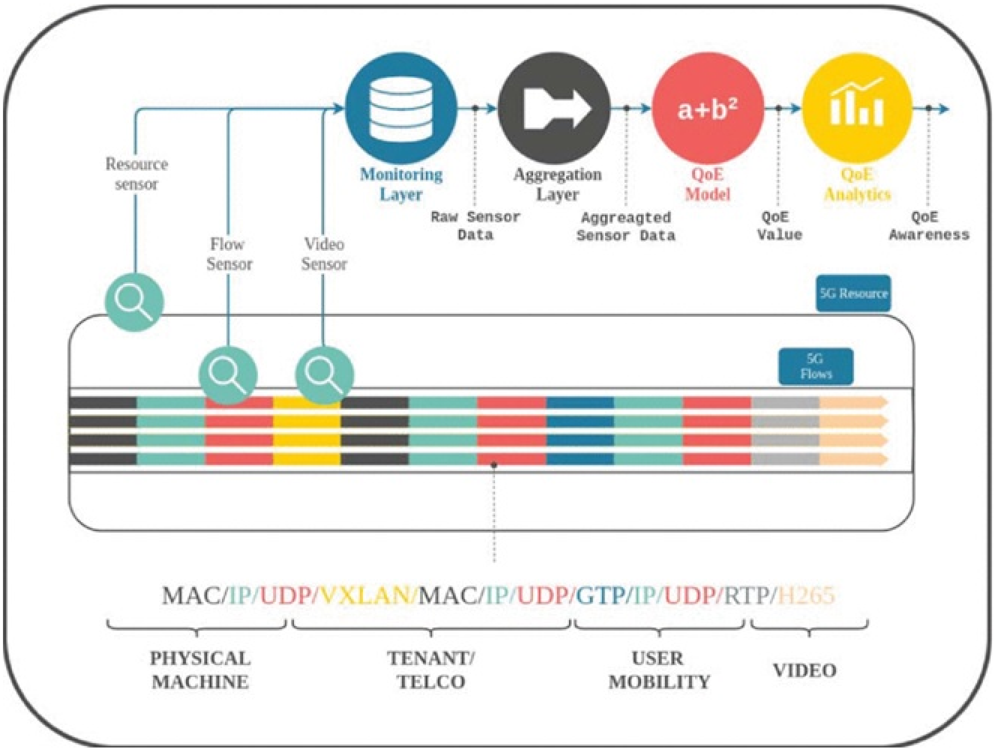


Figure 3.1: Overview of 5G-QoE system architecture

3.1 Network Monitoring Layer

The SELFNET 5G network monitoring layer consists of 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 are stored in a database, used to inform decisions in the SELFNET autonomic network management system and is also available in aggregated form in SELFNET's network management dashboard [**?**].

3.1.1 Flow Sensor

The SELFNET flow sensor is a 5G network agent, which inspects every flow passing through the network, acquiring information and metrics from each level of the complex set of encapsulations. 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 are collected on a configurable periodic basis and pushed to a monitoring database.

3.1.2 Video Sensor

Whenever a new RTP based video flow is detected by the flow sensor, it is immediately mirrored to the video sensor. Supplemental Enhancement Information (SEI) messages are extracted for each scalable layer of the SHVC stream. The SEI messages 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.

3.1.3 Resource Sensor

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.

3.2 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 measures the maximum level of congestion, across all of the network interfaces a flow traverses.

3.3 QoE Modelling

It is to provide a simple, 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. This prediction is known as the video Quality Index. This directly maps to the MOS provided by subjective evaluation experiments. The video Quality Index is SELFNET's primary 'health of network' metric for RTP-based video streaming services.

3.4 QoE Analysis

The 5G-QoE 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.

Methodology And Subjective Testing

4.1 Methodology

This section describes the methodology employed to firstly determine and subsequently validate the proposed 5G-QoE system. Firstly, a set of 4k resolution video clips with varying content types were obtained and encoded in a scalable H.265 format. And 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 network testbed. Fig. 4.1 provides a diagrammatic representation of the methodology and workflow.

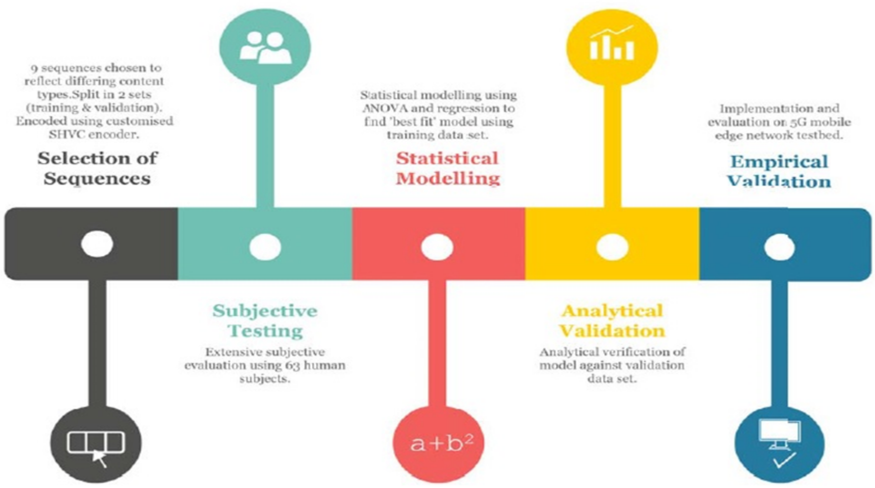


Figure 4.1: Diagrammatic representation of methodology and workflow

4.2 Sequence Selection and Preparation

Nine short video sequences were encoded using a modified version of the scalable HEVC reference software, which inserts additional NAL units containing 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 (eg: Base layer and Enhancement layer). The base layer had a spatial resolution of 1920x1080 Full HD (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 [6] and a second set of 5 sequences each with a frame rate of 24 fps obtained from Ultra Video Group. Fig. 4.2 shows the spatio-temporal characteristics of the employed sequences in terms of the well-known Spatial Index (SI) and Temporal Index (TI). Fig. 4.2 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 I and Table II for the 30 fps and 24 fps sequences respectively.

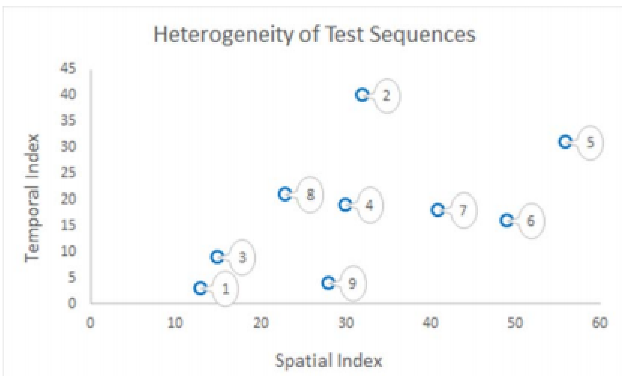
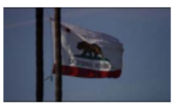


Figure 4.2: Heterogeneity of test sequences

TABLE I
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	

TABLE II
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	

4.3 Subjective Testing Platform

Testing platform (Fig. 4.3) 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 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.

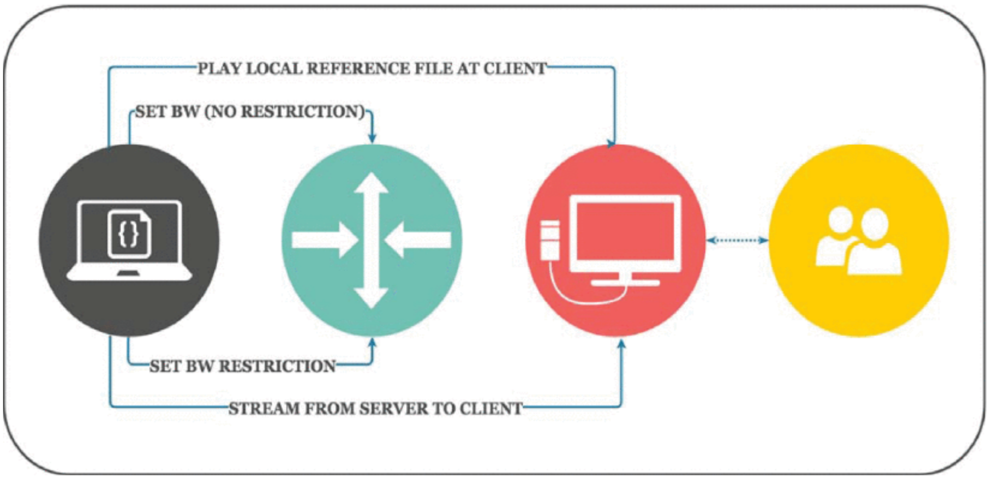


Figure 4.3: Overview of the subjective testing platform

4.4 Subjective Evaluation

Subjective evaluations took place over a period of two weeks. At each testing session viewers sat, in small groups of between five and seven, at the distance recommended in ITU Recommendation BT-500 for the screen size. 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 [7] after a 5 second pause, 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.

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 (ACR) scale shown in below. Viewers had the opportunity to repeat any individual comparison before providing their opinion.

ACR SCALE

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

Initially the bandwidth between server and client devices is unrestricted. The local reference file is played at the client, an appropriate pause 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 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 is repeated for the next test sequence.

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 (Equ: 4.1), where R is the rating provided by an individual user and N is the number of users:

$$MOS = \frac{\sum_{n=1}^{N} R_n}{N} \qquad \text{(Equ: 4.1)}$$

MOS-Mean Opinion Score

R-Rating provided by individual user

N-Number of users

4.5 Subjects

A total of 64 human subjects took part in the subjective evaluations, and they were drawn from both University staff (lecturers and research staff) and the student body (from applied computing disciplines). 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. 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.

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 and then the statistical methods used to derive the candidate QoE index formula. The two validation regimes are then 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) 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 the ratio of the maximum required bandwidth for the stream divided by the available bandwidth for the stream:

$$CI = \frac{M}{A} \qquad \text{(Equ: 5.1)}$$

M-Maximum Required Bandwidth
A-Available Bandwidth
CI-Congestion Index

Statistical analysis of the data using one-way ANOVA, regression and curve fitting produced the function below providing the best fit within the limits of the range of CI examined:

$$Q = -0.891 + \frac{5.082}{\sqrt{\frac{M}{A}}} \quad \text{Q-Quality Index} \qquad \text{(Equ: 5.2)}$$

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 is either very good with little or no perceptual difference. 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).

The upper boundary is at a Congestion Index of 0.8, meaning that the bandwidth available for the video stream is 25% higher than the maximum bitrate of the stream. The lower boundary is set at 1.8 because 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 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.

```
import math
# Set the values of the constants
b = -0.891
a = 5.082
# Set the boundaries and triggers
upper_boundary = 0.8
lower_boundary = 1.8
quality_alert_trigger = 3.0
# calculate the Quality Index
congestion_index, flow_id = db_query_ci()
if congestion_index < upper_boundary:
  quality_index = 5  # sufficient bandwidth to guarantee delivery
elif congestion_index > lower_boundary:
  quality_index = 1  # no prospect of delivery at this bandwidth
else:
  quality_index = b + (a/math.sqrt(congestion_index))
#Limit the range of values
if quality_index > 5:
  quality_index = 5
elif quality_index < 1:
  quality_index = 1
#Raise alert when quality falls below acceptable levels
if quality_index < quality_alert_trigger:
  raise_quality_alert(flow_id, quality_index)
```

Figure 5.1: Pseudo-code for QoE prediction and alert

Comparisons between the predicted QoE of training set and the actual MOS perceived by the human subjects are shown in Table IV, which highlights the correlation, 0.01 significance level, and Fig. 5.3, which provides a graphical comparison of the same data set.

(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).

Figure 5.2: Correlation Between the Predicted QoE and Actual MOS for the Training set

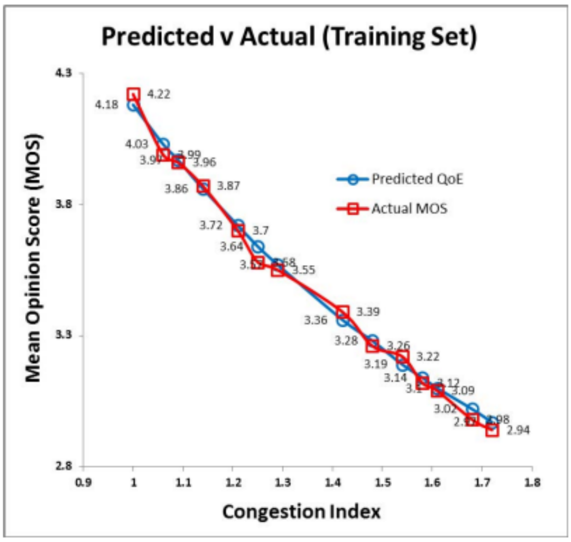


Figure 5.3: Comparison of the Predicted QoE with the Actual MOS in the Training set

QOE Model Validation

6.1 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 sequences in the validation set with the actual Mean Opinion Scores provided by human subjects during the subjective tests. The Congestion Index was calculated.

In Fig. 6.2, 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. From Fig. 6.2, it can be clearly seen that the predicted QoE results closely track the average subjective MOS scores provided by human subjects. It can be seen from Fig. 6.2 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 V that the correlation between the predicted QoE and actual MOS is very high.

TABLE V
CORRELATION BETWEEN PREDICTED QOE AND
ACTUAL MOS FOR THE VALIDATION SET

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

Figure 6.1: Correlation Between Predicted QoE and Actual MOS for the Validation Set

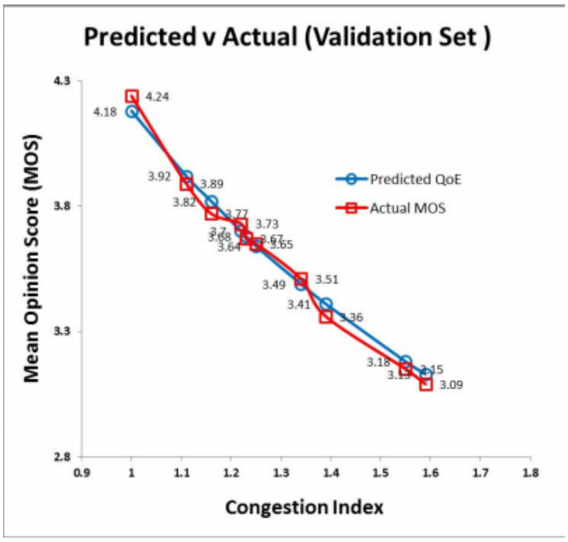


Figure 6.2: Analytical validation by comparing the predicted QoE with the actual MOS in the validation set

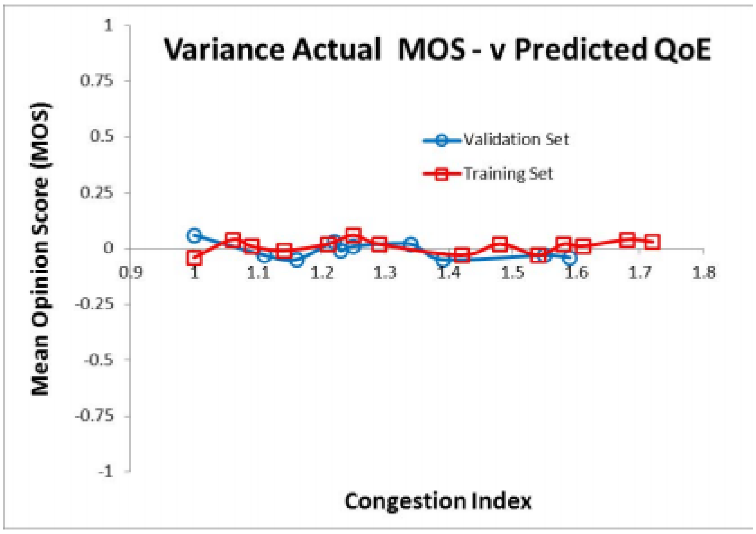


Figure 6.3: Analytical validation by comparing the predicted QoE with the actual MOS in the training and validation sets-–variance

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

TABLE VI
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

Figure 6.4: Comparison of 4K and FHD When CI =1 For The Validation Set

6.2 Emperical Validation

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.

6.2.1 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 testbed was implemented using the open source OpenAirInterface implementation, and the 5G infrastructure has been achieved through introducing the CloudRAN 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, PGW (Packet Data Network Gateway), and SGW (Serving Gateway).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 containing all of the encapsulation layers found in a multi-tenant 5G mobile network infrastructure.

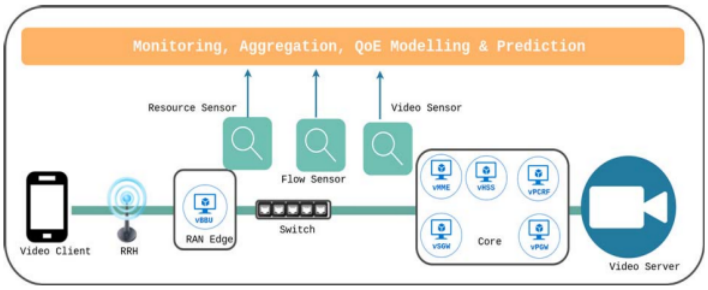


Figure 6.5: 5G Infrastructure Testbed with the positioning of the 5G-QoE system to create and capture realistic 5G traffic

6.2.2 Empirical 5G-QoE Testbed

Fig. 6.6 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 such as the various software sensors, and the monitoring, aggregation, QoE modelling and prediction and analysis software modules.

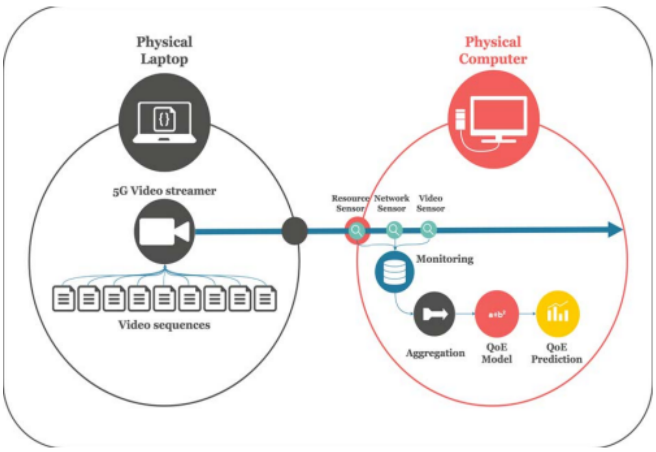


Figure 6.6: 5G-QoE testbed for empirical experiments

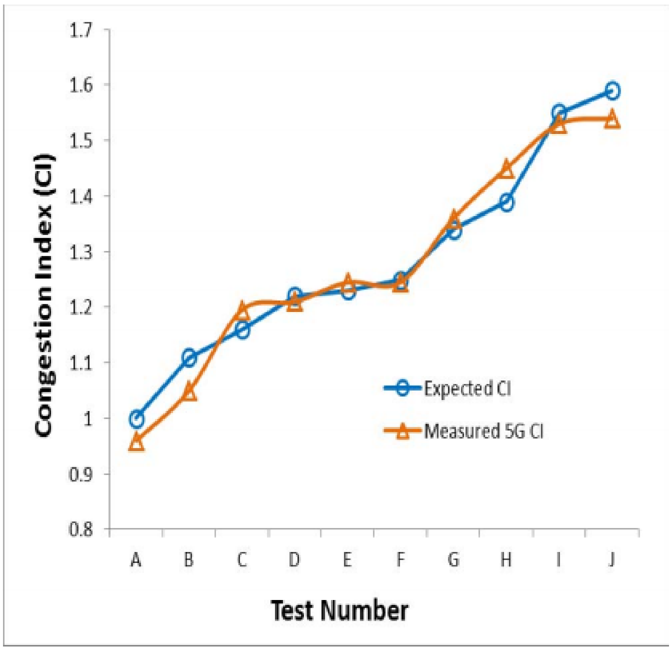


Figure 6.7: Empirical validation by comparing expected CI with measured CI(vs. test number)

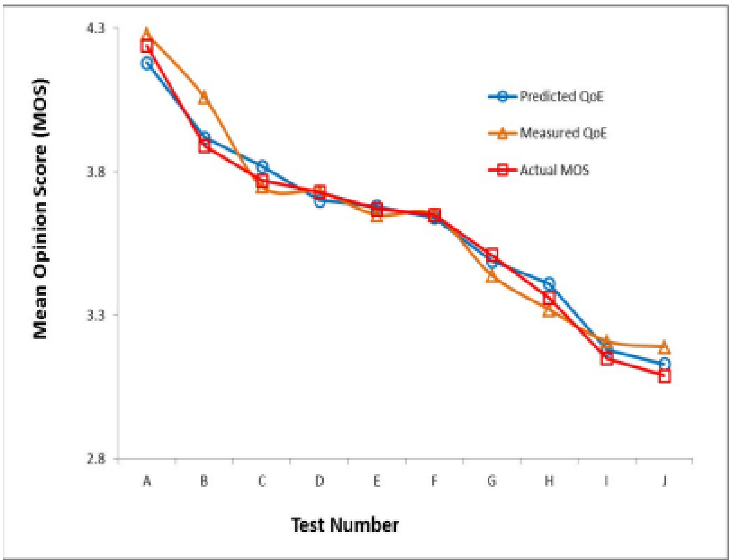


Figure 6.8: Empirical validation by comparing predicted QoE with measured QoE and actual MOS (vs. test number)

The 10 testing points shown in Fig. 6.7 and Fig. 6.8 were average 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) correspond to the letters A to J.

6.2.3 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. 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.

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. 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 means that an average MOS score of 4.3 from our subjective testing indicates a high level of satisfaction with the quality of video.

The components in the 5G-QoE system (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. 6.7 provides a graphical comparison between the expected CI (analytically derived) and the measured CI obtained during empirical validation. Each of the testing points was created by varying the available bandwidth and video sequences, 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.

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

Comparing between actual MOS, predicted QoE and empirically measured QoE, the results shows a strong correlation between each set of results. A comparison of MOS/QoE scores is provided in Fig. 6.8, which shows that scores for empirically measured QoE closely track those for both predicted QoE and the actual MOS scores from subjective testing. Fig. 6.8, which allows easy comparison with Fig. 6.10. Provides a plot of mean opinion score against CI 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. Similar observations can be made when actual MOS and empirically measured QoE are compared.

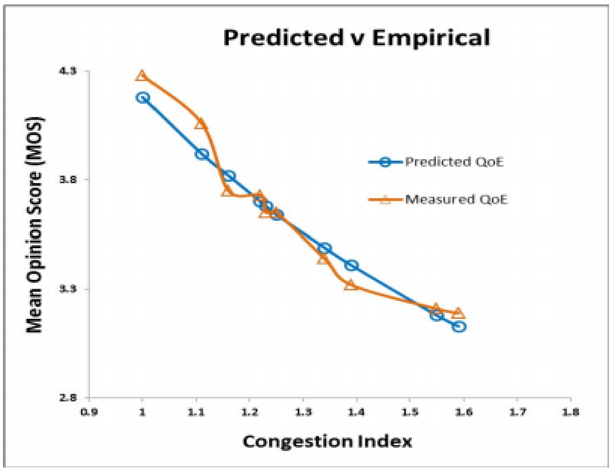


Figure 6.9: Empirical validation by comparing predicted QoE with measured QoE (vs. CI)

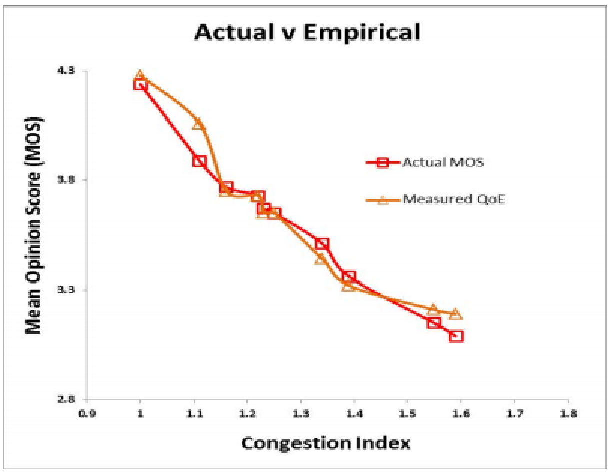


Figure 6.10: Empirical validation by comparing measured QoE with actual MOS (vs. CI)

Conclusion

This study has presented a fast and scalable method of estimating the perceived quality of experience of users of UHD video flows in the 5G networks. 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 of only 0.06 and 0.17 respectively. s. 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.

References

[1] J. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, “Comparison of the coding efficiency of video coding standardsincluding high efficiency video coding (HEVC)”, *IEEE Trans. Circuits Syst. Video Technol*, vol. 22, no. 12, Dec 2017, pp. 1669 -– 1684.

[2] “High efficiency video coding”,” *Int. Telecommun. Union, Geneva, Switzerland, ITU-T Recommendation H.265 (V4),Dec 2016(Online)*

[3] *The EU 5G PPP SELFNET Project. (Apr. 2017). SELFNET: A Framework for Self-Organized Network Management in Virtualized and Software Defined Networks. [Online]. Available: https://selfnet-5g.eu/,*

[4] A. Takahashi, D. Hands, and V. Barriac, ““Standardization activities in the ITU for a QoE assessment of IPTV,” *” IEEE Commun. Mag,* Vol. 46, no.2, pp. 78–84, Feb. 2008

[5] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, “Overview of SHVC: Scalable extensions of the high efficiency video coding standard, ” *IEEE Trans. Circuits Syst. Video Technol.,* Vol. 26, no. 1,pp. 20–34, Jan. 2016.

[6] *Ultra Video Group. (2017). 4K Test Sequences. [Online]. Available:http://ultravideo.cs.tut.fi/#testsequences*

[7] Mitch Martinez. (2017) *.4K Video Sequences. [Online]. Available: http://mitchmartinez.com/free-4k-red-epic-stock-footage/*

[8] “Methodology for the subjective assessment of television pictures,”*Int. Telecommun. Union, Geneva, Switzerland, ITU-Recommendation BT500-13, Jan. 2012*