OPEN ACCESS

UNIVERSITY OF THE
WEST *of* SCOTLAND
UWS

UWS Academic Portal

5G-UHD

Salvá Garcia, Pablo; Alcaraz Calero, Jose M.; Marco Alaez, Ricardo; Chirivella Pérez, Enrique; Nightingale, James; Wang, Qi

Link to publication on the UWS Academic Portal

Download date: 17 Sep 2019

Accepted Manuscript

5G-UHD: Design, Prototyping and Empirical Evaluation of Adaptive Ultra-High-Definition Video Streaming Based on Scalable H.265 in Virtualised 5G Networks

Pablo Salva-Garcia , Jose M. Alcaraz-Calero ,
Ricardo Marco Alaez , Enrique Chirivella-Perez ,
James Nightingale , Qi Wang

Please cite this article as: Pablo Salva-Garcia , Jose M. Alcaraz-Calero , Ricardo Marco Alaez , Enrique Chirivella-Perez , James Nightingale , Qi Wang , 5G-UHD: Design, Prototyping and Empirical Evaluation of Adaptive Ultra-High-Definition Video Streaming Based on Scalable H.265 in Virtualised 5G Networks, *Computer Communications* (2017), doi: 10.1016/j.comcom.2017.11.007

5G-UHD: Design, Prototyping and Empirical Evaluation of Adaptive Ultra-High-Definition Video Streaming Based on Scalable H.265 in Virtualised 5G Networks

Pablo Salva-Garcia, Jose M. Alcaraz-Calero, Ricardo Marco Alaez, Enrique Chirivella-Perez, James Nightingale, Qi Wang*

Email addresses: Pablo.Salva-Garcia@uws.ac.uk, Jose.Alcaraz-Calero@uws.ac.uk, Ricardo.MarcoAlaez@uws.ac.uk, Enrique.Chirivella-Perez@uws.ac.uk, James.Nightingale@uws.ac.uk, Qi.Wang@uws.ac.uk.

*Corresponding author. Tel.: +441418483945

School of Engineering & Computing, University of the West of Scotland, Paisley PA1 2BE, United Kingdom

ABSTRACT

Ultra-High-Definition (UHD) video applications such as streaming are envisioned as a main driver for the emerging Fifth Generation (5G) mobile networks being developed worldwide. This paper focuses on addressing a major technical challenge in meeting UHD users' growing expectation for continuous high-quality video delivery in 5G hotspots where congestion is commonplace to occur. A novel 5G-UHD framework is proposed towards achieving adaptive video streaming in this demanding scenario to pave the way for self-optimisation oriented 5G UHD streaming. The architectural design and the video stream optimisation mechanism are described, and the system is prototyped based on a realistic virtualised 5G testbed. Empirical experiments validate the design of the framework and yield a set of insightful performance evaluation results.

1. Introduction

The next-generation, Fifth Generation (5G), mobile networks are emerging from the horizon at an accelerated rate thanks to the intensified global actions on 5G research, development and standardisation in the last few years. In Europe, the European Union (EU) Commission and 5G stakeholders such as telecommunication operators, vendors, service providers, Small and Medium-sized Enterprises (SMEs) and academia have initiated the 5G Infrastructure Public Private Partnership (5G PPP) to advance 5G strategic development in Europe and beyond, in collaboration with other major global 5G initiatives [1]. 5G PPP is coordinating 19 Phase 1 5G projects [2], which were launched in 2015 and cover a full range of key 5G technological areas from new radio spectrum to improved user-facing services to fulfil a set of 5G Key Performance Indicators (KPIs) (e.g., [3][4]) and visions (e.g., [5][6]).

Among these 19 ongoing 5G PPP projects, the SELFNET project [7][8] focuses on cognitive network management taking a self-organising approach for intelligently managing virtualised 5G networks and services. A primary use case of SELFNET is self-optimisation of Ultra-High-

Definition (UHD) video streaming in 5G hotspots. A 5G hotspot is a communication hotspot covered by a 5G network, comparable to its WiFi counterpart, and examples include airport, campus, stadium, cafe and other public or private venues. This use case handles delivering high-quality video streams from a network media server to end users in a 5G hotspot with sustained users' perceived quality, referred to as Quality of Experience (QoE). In this 5G hotspot, a number of end users are actively engaged in high-quality video streaming applications, potentially UHD videos as foreseen as a main driving application in 5G networks. In fact, video applications especially HD and UHD services have been gaining growing popularity and consequently video traffic has dominated the bandwidth of Internet and mobile networks. In 2012, mobile video traffic exceeded 50% for the first time in history. This domination is ever-increasing: it is forecasted that video traffic will account for nearly 75% of global traffic by 2019 [9]. In this context, congestion in the network is likely to happen, which would negatively affect end users' QoE. This use case scenario is designed to demonstrate the self-optimisation capability of SELFNET in maintaining or minimising the disruption of end users' QoE under this adverse circumstance. To help reduce the sheer volume of video traffic, advanced video coding standards have been taken into account. The High Efficiency Video Coding (HEVC) [10] or known as the H.265 video coding standard [11] published by the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) has been proven to reduce up to 50% of bandwidth requirement of a video flow compared with its precursor, H.264. Consequently, H.265 is expected to be widely adopted by wired and wireless networks in the next years, in line with the deployment of 5G networks, e.g., as highlighted by the Next Generation Mobile Networks (NGMN) Alliance [12]. This use case scenario further advances this vision by leveraging the scalable extension of H.265, known as Scalable H.265 or Scalable HEVC (SHVC) [13][14], for QoE-informed video streaming adaptation in congestion-prone 5G areas. The scalability of this latest standard codec in multiple dimensions (spatial, temporal, quality and so on) can be exploited to realise efficient video adaptation based optimisation. However, little existing work has been reported in the public domain on Scalable H.265 based HD/UHD streaming over a 5G network.

Achieving a novel system that enables the above use case would help contribute to realising a number of 5G KPIs [3][4] and visions [5][6]. In particular, the technologies developed can enable or improve new-generation economically-viable UHD-based services of high societal value such as UHD TV. Moreover, the self-optimisation capability in terms of QoE-aware UHD video adaption can pro-actively react to predicated congestion and thus minimise service disruption towards achieving "zero perceived downtime" for service provision. Furthermore, the bandwidth resources freed up by adopting the latest coding standard can be utilised to accommodate more users in the system or other services and applications, thereby leading to greater revenue for the network operators and service providers as well as improve user experience in terms of network admission control given the constraint of network capacity.

The main contributions of this paper are multifold:
- Firstly, this research proposes and prototypes a new adaptive 5G-UHD framework towards achieving self-optimising UHD video streaming in emerging 5G networks. Both design and implementation details are provided, filling a gap in the literature on this topic. In particular, a practical multi-purpose 5G-aware traffic actuator is emphasised for the purpose of enforcing the video self-optimisation actions.
- Secondly, this research explores the latest video codec, Scalable H.265, and implements

real-time streaming and optimisation based on Scalable H.265. It is noted that H.265 and Scalable H.265 manage to achieve significantly higher compression efficiency compared with H.264 at the cost of substantially increased complexity and computational power. It is therefore challenging to prototype Scalable H.265 streaming and optimisation in a virtualised, commodity out of shelf equipment based 5G testbed.

- Lastly, this research provides empirical evaluation of the performance of the proposed 5G-UHD system over realistic virtualised 5G infrastructure. These numerical results lead to insights on the performance of video adaptation and the impact of some of the new challenges posed by the introduction of key enablers of 5G including softwarisation, virtualisation and multi-tenancy.

The organisation of the remainder of the paper is as follows. Section 2 reviews related work. Section 3 presents the proposed 5G-UHD framework highlighting the design of the system architecture and the central component for video adaptation/optimisation. Section 4 describes the prototyping of the system and the validation testing and results, and Section 5 provides further details on performance evaluation of the system. Section 6 concludes this paper and outlines the future work.

2. Related work

2.1 5G networking technologies and video application requirements in 5G networks

5G research and development efforts have gained momentum in the last few years. Upon completing the first stage's tasks, the 19 ongoing EU 5G PPP projects have recently reported 15 Phase 1 Golden Nuggets [15] that reflect the latest 5G advances in Europe. These include the communication and networking aspect such as 5G Spectrum Requirements, Evaluation and Candidate Bands, 5G Massive Channel Access, 5G Flexible Radio Access Network (RAN), 5G Integrated Transport Networks, as well as the service and network management aspect such as Flexible and Agile Service Deployment, 5G Network Management, E2E Orchestration in Single and Multi-Domains 5G Virtualized Networks, 5G Networks Security and Integrity etc.

In addition, 5G is widely envisioned to be established upon a number of cornerstone technologies [16] including cloud computing and network virtualisation, Network Function Virtualisation (NFV), softwarisation such as Software-Defined Networking (SDN) and the enabled programmability, and multi-tenancy and thus network sharing among network operators. These 5G enablers are also highlighted as 5G PPP Phase 1 Golden Nuggets such as Network Softwarisation and Programmability Integrating SDN and NFV Technologies, and Programmable Industrial Networks. This softwarisation based paradigm shift in 5G will allow significantly enlarged service innovation, accelerated service time-to-market, and strengthened network agility, among others, all at reduced Capital Expenditure (CAPEX) by employing Commodity-Off-The-Shelf (COTS) equipment and Operational Expenditure (OPEX) through software-defined, centralised management, among others. Meanwhile, the softwarisation and virtualisation may also introduce overhead and performance constraints to services and applications especially resource-demanding applications such as UHD streaming. However, the impact of this has not been sufficiently studied before; particularly this has not been empirically evaluated in a Scalable H.265 based 5G UHD streaming use case in the public domain.

Although the H.264 family codecs still dominate the current video coding and streaming market,

the new-generation codecs H.265 and Scalable H.265 are expected to replace their H.264 counterparts soon and play the leading role in 5G video applications. The NGMN Alliance has recently identified the technical requirements for a number of selected video application use cases, among other use cases, across the vertical industry and made recommendations on codecs to make these use cases economically viable [12]. For instance, for automotive use cases, H.265 HD stream at about 10 Mbps is suggested for applications such as dynamic high-definition digital map update, video sensor sharing and remote drone operation. For more demanding video services such as eHealth and mHealth use cases, H.265 HD at about 10 Mbps is recommended to provide the basic service, whilst UHD 4K at 30~40 Mbps or even 8K at 80~100 Mbps are desired for premium services. These recommendations, however, need to be further investigated and optimised for specific use case scenarios such as the one reported in this paper when the dynamic network conditions (e.g., congestion) are taken into account as constraints.

2.2 Scalable H.265 HD/UHD video streaming in 5G networks

As highlighted, Scalable H.265 ([13], [14]) is the latest standard video codec. A Scalable H.265 video stream consists of multiple layered video flows, where a base layer comprises the essential yet basic representation of the video whilst one or more enhancement layers contain additional details to enhance the base layer to make the composite video stream to be scalable in one or more domains. H.265/HEVC includes built-in temporal scalability, whilst Scalable H.265 further defines extension tools to provide spatial, quality/fidelity (SNR), bit depth, and colour (gamut) scalability for H.265/HEVC. However, research on Scalable H.265 based HD/UHD streaming in any networks is in its infancy. Existing work on H.265 or Scalable H.265 still focus on improving signal processing in the encoder (e.g., [24]) or the decoder (e.g., [25]), rather than on streaming with the exception of some preliminary work reported in a few conference papers on rate control [31], rate-distortion performance [32] or demo [33], respectively.

Indeed, there is a clear lack of academic publications on UHD streaming, whether H.265 (let alone Scalable H.265) based or not, over 5G networks. Mainly relevant industry trials hit the headlines. For example, Ericsson and Sprint showed a 4K UHD Video Streaming demo in a 5G network in 2016 [26], and Nokia and DoCoMo managed a 5G 8K demo in the same year [27]. However, no technical details such as the employed codec are reported. It is also noticed that the purpose of these industry trials emphasise the increased speed of 5G networks and thus the UHD streaming trials in most times if not all were performed in over-provisioned bandwidth conditions. In contrast, 5G-UHD focuses on UHD streaming in congestion-prone conditions. Among other reasons, the telecommunications history has indicated that the growth of users' requirements could soon outpace that of a mobile system's capacity.

Meanwhile, the UHD-on-5G project [28], which started in 2016, is one of the very few related projects that has produced academic publications. The focus of the project is to explore Information-Centric Networking (ICN) and SDN technologies for efficient 5G UHD streaming. In their publications so far e.g., [29], no standard or Scalable H.265 based UHD streaming is highlighted though. In fact, almost all existing publications on UHD video streaming are either not delivered in 5G networks or has explored Scalable H.265. Most studies adopt the precursors of H.265, especially H.264 or its scalable extension Scalable Video Coding (SVC) [22] for streaming (e.g., [23]).

It is noted that H.265 can be utilised in either Hypertext Transfer Protocol (HTTP) or Real-time

Transport Protocol (RTP)/User Datagram Protocol (UDP) based streaming. This also applies to Scalable H.265. There are advantages and disadvantages in either of these two main streaming approaches [30]. For instance, HTTP based streaming especially those using the Dynamic Adaptive Streaming Over HTTP (DASH) standard [19] defined by the Moving Picture Experts Group (MPEG) has been gaining popularity as it has less difficulty in traversing firewalls whilst this approach often suffers from annoying stalling and initial playback delays, which degrade the user's QoE [20]. On the other hand, RTP/UDP based streaming produces smoother real-time playback experience and yields smaller overhead yet at the cost of picture fidelity in lossy networks and can be blocked by firewalls. In the prototyping work reported in this paper, the RTP/UDP approach is followed for a faster prototyping, following the H.265 RTP packetisation standard defined in RFC 7798 [21] by the Internet Engineering Task Force (IETF). The real-time attribute in this paper thus refers to the delivery of video packets over the network based on the RTP/UDP stack. Nevertheless, it is worth noting that the design of the proposed 5G-UHD framework is applicable to HTTP based streaming too, and will be explored in the future work.

In addition, the Network Abstraction Layer (NAL) unit design in H.265 and Scalable H.265 standards facilitates Media-Aware Network Elements (MANEs) to extract video-related parameters/metadata from the NAL headers and/or signalling NAL units so that further processing such as adaptation can be enabled [46]. For instance, in Scalable H.265, the information on the base layer and enhancement layers is available this way, and thus a Scalable H.265-aware MANE can be designed to process the layers for adaptation/optimisation purposes. This facility is leveraged by the proposed 5G-UHD framework in this paper. Little existing work has reported any details on MANE design or prototyping for Scalable H.265 UHD applications. It is noted that the Third Generation Partnership Project (3GPP) has indicated in TR 26.906 [47] that the temporal scalability embedded in H.265 can be explored by an H.265-aware MANE for adaptation although no detailed recommendation was provided.

3. The proposed 5G-UHD framework
This section describes in details the architectural and central component design of the 5G-UHD framework.

3.1 Architectural design
The system architecture of the proposed 5G-UHD framework is illustrated in Figure 1. The framework is built upon a reference 5G infrastructure consisting of RANs, a core network, and Internet. For illustration and prototyping purposes, the networking terminologies used in the Fourth Generation (4G) mobile network, the Long Term Evolution (LTE), are adopted here. It is however noted that the framework is applicable to further evolved 5G networks that employ new air interface, RAN and core network. This reference 5G infrastructure is constructed through evolving and virtualising the correspondent LTE components, following the network softwarisation/cloudification approach in 5G. In line with the Multi-access Edge Computing (MEC) paradigm [40] envisioned by the European Telecommunications Standards Institute (ETSI), the Edges of the infrastructure employ the Cloud-RAN model, where BaseBand Unit (BBU) and Remote Radio Head (RRH) functions are separated from the conventional LTE eNodeB, and a Data centre hosts the virtualised LTE's core network components including Mobility Management Entity (MME), Home Subscriber Server (HSS), Serving Gateway (SGW), Packet Data Network Gateway (PGW), and Policy and Charging Rules Function (PCRF, not shown as not prototyped yet in this work) in the format of Virtualised Network Functions (VNFs).

In this 5G infrastructure, our 5G-UHD contribution proposes and deploys a new component called 5G Flow Control Agent (FCA). A 5G FCA is a network actuator capable of processing data plane traffic of any kinds and video streams/flows is not an exception. This functionality can be used to achieve video adaptation/optimisation (please refer to Section 3.2 for more design details). FCAs can be deployed collocated with other system components such as the Edge Services server in a RAN, or deployed on demand to be placed in any other segment in the infrastructure. These distributed FCAs report and listen for actions to/from a shared communication middleware where an FCA Manager component can collect the reported messages and configure FCAs for network management and control purposes.
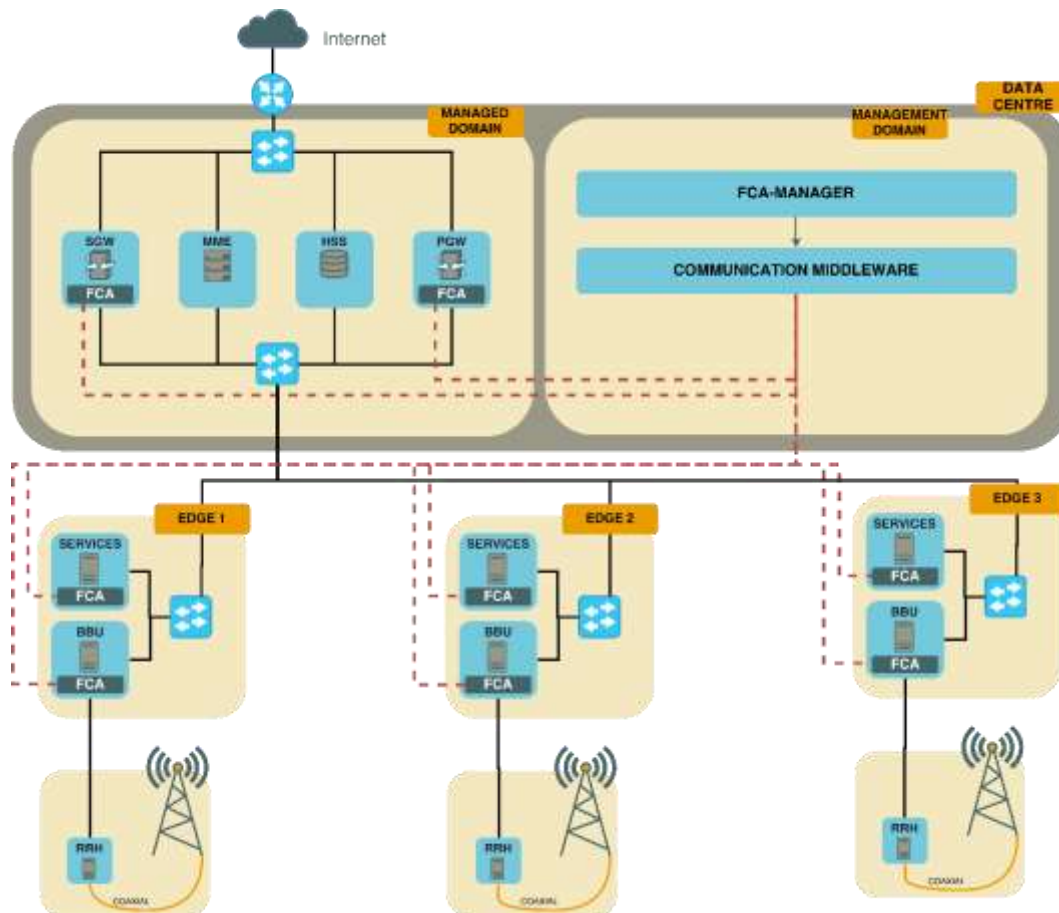

Figure 1. The proposed 5G-UHD system architecture. Edges are interconnected to a Data centre with 5G-LTE components virtualised, and FCAs are distributed to any interested locations.

3.2 Video flow processing design

3.2.1 5G FCA-enabled flow traffic processing

The 5G FCA is designed to be a multi-purpose tool in order to have control over the 5G data plane. It enables different video flow adaptation/optimisation schemes based on traffic engineering. Consequently, an FCA can be configured on demand to act as a Media-Aware Network Element (MANE) to enforce video traffic adaptation (through selective packet dropping), a Traffic Redirection server to reroute traffic, or a Traffic Mirroring server to mirror traffic for monitoring purposes, based on leveraging and extending the Linux kernel Netfilter module [37]. Figure 2 shows the design details of FCA, including its building blocks and interfacing with the communication middleware and underlying system components at the kernel level, as described below.
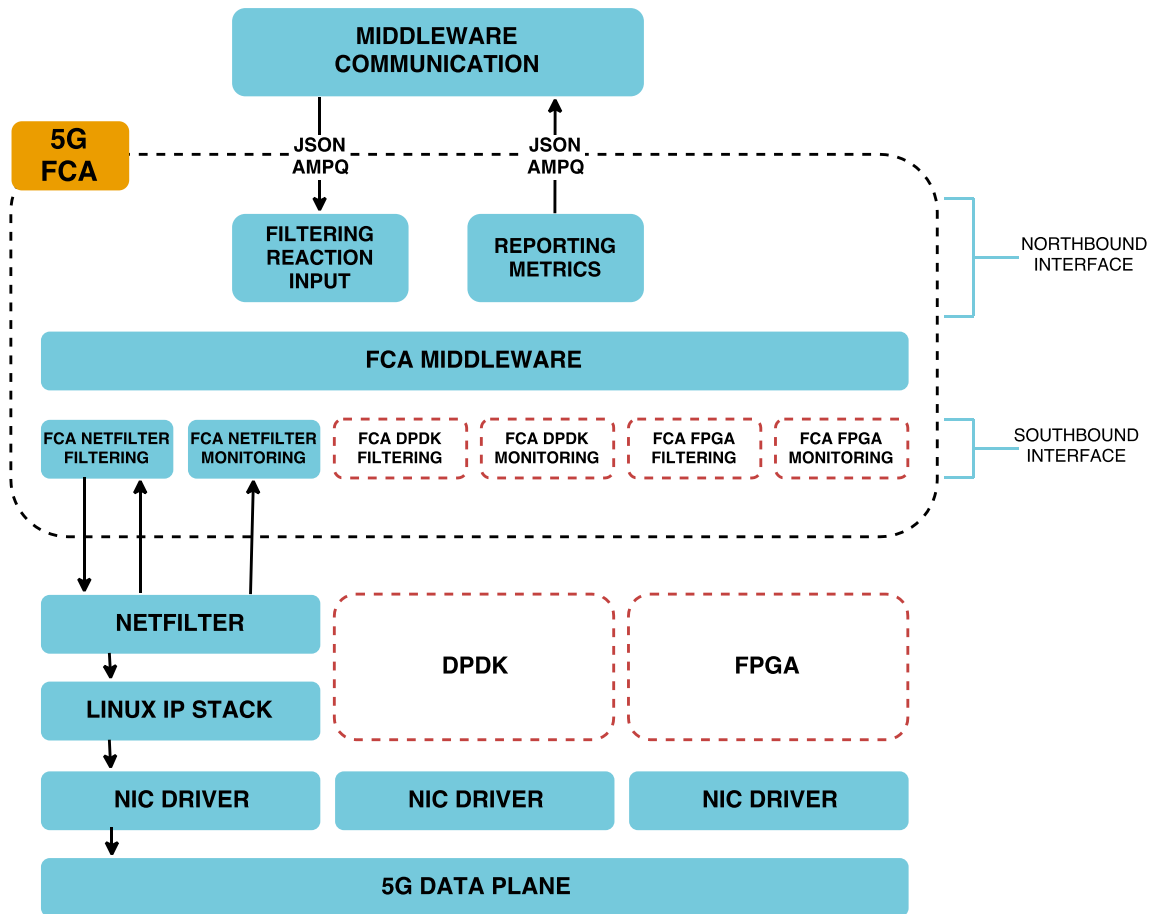
Figure 2. Design of the video traffic optimiser (FCA). A layered approach is followed with essential FCA components prototyped and alternative components indicated (dotted lines).

The architecture of the FCA is divided in northbound and southbound interfaces. The northbound interface is used to receive and send information from/to a communication middleware such as RabbitMQ [44] and Kafka [45], among others. The southbound interface is used to configure filtering capabilities of the data plane. The northbound interface is composed by an output interfaces for reporting metrics and an input interface to receive data plane configurations as reactions.

Filtering Reactions Interface: This interface is used to configure the filter to be enforced in the data plane. The interface enables the definition of a filter applied over a Scalable H265 video

flow being transmitted over a 5G network. The definition of this interface is a key contribution of this paper. The communication protocol is based on the JavaScript Object Notation (JSON) [52] and Advanced Message Queuing Protocol (AMQP) [53]. Listing 1 illustrates a JSON example of the data model received in this interface to perform a video adaptation by selectively dropping any layer upper than layer ID 1 of an Scalable H.265 video flow.

```
{
    "Reaction": {
        "Location": "146.191.50.1",        #FCA IP: where to apply the action
        "Action": "NEW",                   #Create a NEW Filtering Rule
        "Type": "DROP"                     #Filtering Action = Drop
    },
    "Flows": [
        {
                "encapsulationLayer": 2,    # Number of tunnels
                "flowLayer": 3,             # DPI Functions (2 for tunnels + 1 for video)

                "l7Proto": "rtp",           #Classified Traffic
                "srcIP": "192.188.0.140",   #Streamer PC IP
                "destIP": "192.188.0.2",    #Client 5G Mobile IP
                "l4Proto": "17",            #Transport Protocol UDP
                "srcPort": "5000",          #Streamer PC Port
                "destPort": "5000",         #Client 5G Mobile Port
                "l7Key1": "96",             #Dynamic RTP Payload
                "l7Key2": "1",              #SHVC Layer ID
        }
    ]
}
```

Listing 1. JSON example of an FCA filtering reaction.

Reporting Metrics Interface: This interface provides metrics about the amount of traffic being dropped by the rules that have been installed in the data plane. It allows the system to know how much traffic has been selectively dropped and how much bandwidth saving has been made through this adaptation. This interface was used in the experiments described in Section 5 to gather the statistics on packets intentionally dropped and to be able to differentiate these packets from random packet losses in network.

FCA Middleware: This is the core business logic of the FCA components. It is in charge of receiving reactions from the filtering reconfiguration interface and to apply such configurations in the selected southbound implementation for filtering packets in the data plane. Essentially, it will produce a different reaction implementation depending on filtering selected. The FCA middleware will report metrics per every reaction enforced.

FCA southbound interface is designed using a modular pluggable approach. Different implementations can be attached to the southbound interface. Each of the implementations should provide at least two different APIs: filtering and monitoring.

FCA Southbound Filtering Interface: The FCA has been designed to provide modular filtering options such as Netfilter [37], Data Plane Development Kit (DPDK) [48] or NetFPGA [49] based on the Field-Programmable Gate Array (FPGA), among others. In this paper, Netfilter has been implemented as part of the prototyping. Netfilter allows using a set of hooks inside the Linux kernel to provide callback functions for every packet that traverses the respective hook within the network stack. Netfilter has a table structure for the definition of rulesets, known as iptables, which allows defining classifier and action methods over the network traffic. The FCA Netfilter

Filtering module uses these Netfilter tools in order to enforce rules therein.

FCA Southbound Monitoring Interface: The FCA allows a pluggable system to attach different monitoring approach to be able to retrieve different metrics about the enforced rules into the data plane. It allows selecting an implementation among different options. A common reporting structure is sent to the FCA middleware. The included metrics indicate the measured network performance resulted from the applied rules.

5G Data Plane: This represents packets flowing path through the network to this host. The main innovation is that these packets are not traditional video over IP traffic only but video over novel virtualized 5G traffic. This change has been a motivation of this contribution.
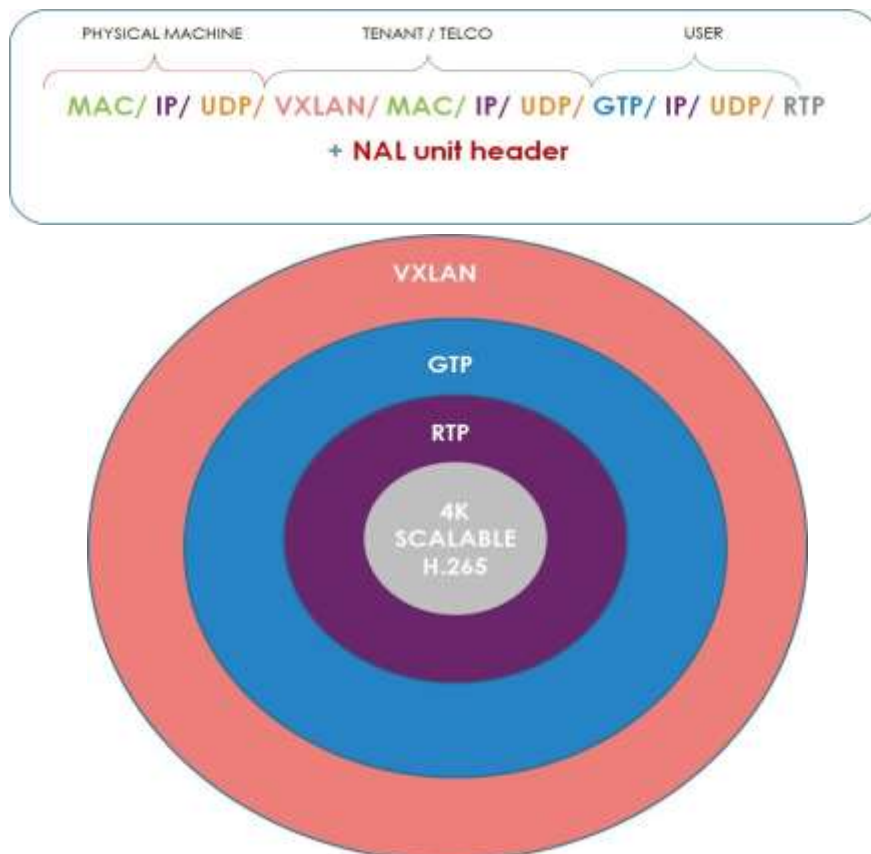

Figure 3. Hierarchical encapsulation for a video packet over a 5G network. The different header groups from the different protocol layers in the data plane are highlighted.

Figure 3 shows the hierarchical encapsulation for an RTP/UDP-based Scalable H.265 video packet being transmitted over a 5G data plane. As shown, these headers can be classified into several groups following the hierarchy. The first group of headers including MAC/IP/UDP is related to the communication between physical machines. The second group including VXLAN/Medium Access Control (MAC)/IP/UDP is inserted to isolate tenant traffic, especially a telecommunication operator sharing the same physical 5G infrastructure as a tenant. VXLAN [36] is a network virtualization technology, and it is employed to allow VLAN-IDs to be re-used and applied to distinguish different tenants so that the multi-tenancy requirement by 5G physical infrastructure sharing can be fulfilled. The next group including General Packet Radio Service (GPRS) Tunnelling Protocol (GTP)/IP/UDP/RTP is used to allow end user's mobility. GTP is the tunnelling protocol employed in LTE-based infrastructures to establish the data plane

for end users with features such as QoS, mobility, admission control, etc.. In addition, the NAL unit is the RTP payload, which stores the video control information and content in this NAL unit's header and payload respectively.

In the scenario where the same video packet is being delivered over a pure IP network, only a subset of these headers will be needed, including MAC/IP/UDP/RTP/NAL. Compared with the pure IP case, several additional headers have been added to achieve multi-tenancy-enabled transactions over a tunnelled infrastructure in a 5G network. The above design needs to be validated empirically and the impact of the additional headers as overhead on performance need to be investigated through empirical experiments as well. Sections 4 and 5 provide this validation together with empirical results.

3.2.2 QoE-informed design for media-aware video adaptation

The media-aware video adaptation scheme adopted is spatial scalability based adaption. It is noted that the design of this scheme has been informed by subjective tests, where viewers of a range of diverse video sequences/clips were asked to give Mean Opinion Scores (MOS) to each of the sequences, in HD and 4K UHD formats respectively at 30 frames per second (default frame rate in all experiments). It is environed that the primary user equipment in 5G is still mobile phones, and thus a smartphone (Sony Xperia Z5 Premium) with a 5.5-inch 4K resolution screen was employed in this research. The MOS testing setup followed the ITU-R BT.500-13 standard [51]. The MOS assessment employed the Double Stimulus Impairment Scale (DSIS) method specified in this standard. The same viewers were presented with a reference 4K UHD video, and then with the HD version of that video immediately. For each pair, each viewer gave a MOS score on the second, ranging from 5 to 0 (5: imperceptible; 4: perceptible, but not annoying; 3: slightly annoying; 2: annoying; 1: very annoying). Then another pair of a different video followed and scored, and so on. The testing results show that 92%~95% viewers indicated that there was no difference between the HD and the UHD versions of the same video sequence on the smartphone's display. The conclusion of the tests is that massive bandwidth saving can be achieved by reducing the spatial resolution from 4K UHD to HD, with practically no impact on the user's perceived quality (QoE). Therefore, under a circumstance of congestion, a 4K UHD video stream can be adapted to an HD one to minimise the QoE downgrading whilst helping mitigate the congestion. Figure 4 shows the 4K video sequences (available at [41]) used in this pre-prototyping test (Beauty, ShakeNDry and ReadySetGo) and in the post-prototyping tests including validation tests (using a concatenated video following the order: Bosphorus, ReadySetGo, ShakeNDry, ReadySetGo, HoneyBee and Bosphorus) presented in Section 4 and performance tests (HoneyBee) in Section 5.
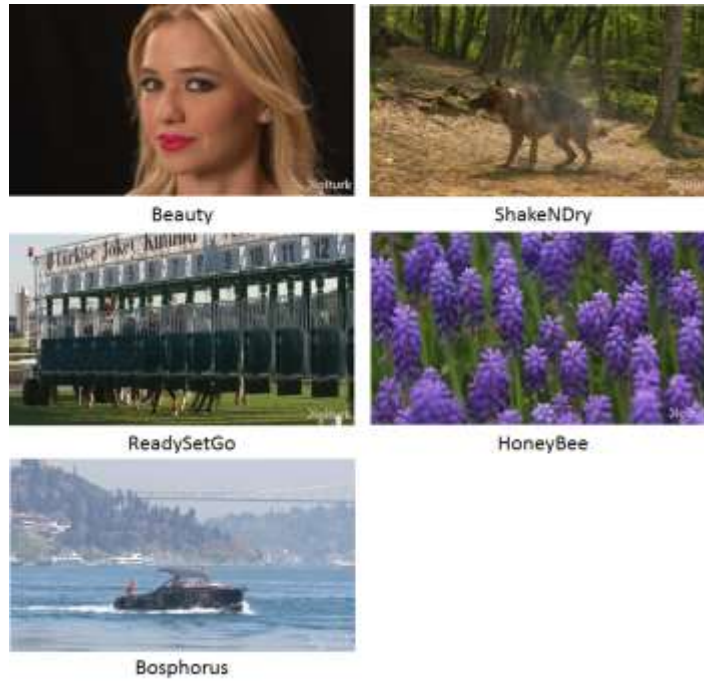
Figure 4. 4K UHD video sequences employed in different tests. These sequences cover a wide range of video scenes of varying characters in objects, motion, texture etc.

4. 5G-UHD prototyping and validation

4.1   Prototyping testbed

To prototype and validate the design of the 5G-UHD system, a testbed as shown in Figure 5, has been implemented. There are two main approaches in developing 5G networks as defined by 3GPP and shown in ongoing 5G projects. One is to develop new air interfaces ("New Radio") different from that of LTE, and the other is based on the evolution of LTE. This research has been implemented over a realistic LTE-based 5G testbed employing OpenAirInterface, a primary platform employed in EU 5G-PPP projects such as SELFNET and Coherent. A main Key Performance Indicator in 5G is reduced CAPEX and OPEX. Therefore, virtualisation and cloudification technologies as the main enablers are commonly utilised to benefit from these technologies primarily for reduced costs for 5G operators.

In compliance with the framework design shown in Figure 1, the testbed comprises the following components/segments. Firstly, the virtualized/cloudified 5G infrastructure is constructed following an MEC paradigm by employing and extending the open-source OpenAirInterface (OAI) platform [34] to achieve a realistic LTE-based testbed. To this end, a Cloud-RAN for an Edge is implemented, where the RRH and the BBU are physically split and the RRH is connected to an LTE antenna, which employs 2 x 2 MIMO and operates at Band 7 (2600 MHz band) of LTE in the Frequency Division Duplex (FDD) mode. The core network, known as the Evolved Packet Core (EPC) in LTE's terminology, is implemented with a set of essential components including an MME, an HSS, an SGW and a PGW. These EPC components, together with the BBU, are implemented as NFVs. OpenStack Mitaka [35] has been deployed as the Virtual Infrastructure Manager (VIM). OpenStack Neutron was installed and controlled by an OpenDayLight Beryllium SDN controller [42], and VXLAN was utilised as the tunnelling technology for isolating tenants.

As to the communication connectivity, the LTE protocol stack from OAI is fully operational to establish both the LTE control and data planes end to end from a video client, known as a User Equipment (UE) in LTE, to a video streamer. The UE is a laptop equipped with an LTE USB dongle (Huawei E398) to gain wireless connectivity with the LTE antenna. The laptop is further connected to a 4K TV via 4K HDMI to show the video playback.

For the UHD video streaming, the standard reference Scalable H.265 encoder SHM4.1 [38] has been used to encode video sequences with spatial scalability from HD (based layer) to 4K UHD (enhancement layer). GPAC 0.6.0 [39] has been utilized as the streamer and client. GPAC software use OpenHEVC as a decoder of the Scalable H.265 video flow in the backend. The video used for all the experiments comprised the concatenated sequences listed in Section 3.2.2.

To enable the corresponding display of HD or UHD video on the client's screen, the prototyping work extended the GPAC client so that it is able to properly respond to the spatial video adaptation (from 4K UHD to HD or vice versa). The network congestion condition can be dynamically emulated and switched on or off by creating or removing a bandwidth bottleneck along the end-to-end route respectively.

Figure 5. 5G-UHD prototyping and validation testbed. This lab-based 5G testbed is based on OAI, employing real LTE-based radio access, Cloud-RAN, and virtualised LTE-based core network.

4.2 *Validation test*

Once the prototyping testbed was established as described above, intensive validation tests were carried out in this project to validate the framework design and the FCA's functionality focusing on its most challenging role as a MANE in adapting real-time Scalable H.265 UHD videos in response to network congestion. Figure 6 shows the screenshot of the I/O graph of Wireshark [50], which was installed at the client to monitor the received video traffic volume.
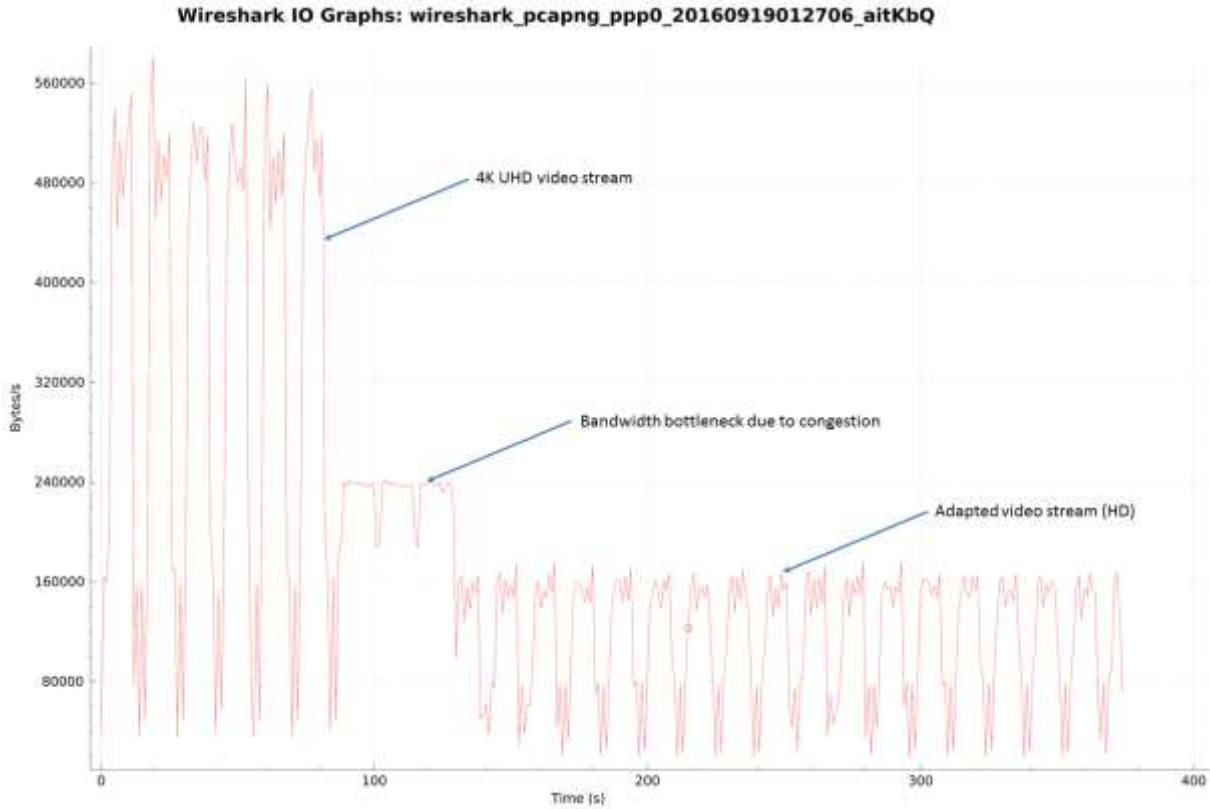

Figure 6. Scalable H.265 video adaptation: validation test result. This Wireshark plot shows the differences in throughput (bandwidth consumption) of the video being streamed in the different phases: original 4K UHD phase, congestion phase, and HD phase (adaption).

As can be observed, one 4K UHD video stream was delivered when the network is not congested, and the UHD stream was later adapted to an HD flow when a network congestion took place. The bandwidth saving is significant through the FCA's selective dropping of the enhancement layer of the Scalable H.265 video. The real-time statistics of the packet dropping behaviours can also be monitored in iptables, which has further confirmed that the FCA works as designed. For the perceived quality on either the 15-inch laptop screen or the 55-inch 4K TV's display, numerous viewers have indicated that little (annoying) differences were noticed during the tests. Therefore, the validation tests were concluded successfully.

5. 5G-UHD performance evaluation

The validation setup presented in Section 4 has allowed the project to gather the Packet Capture (PCAP) file of the communication carried out between the client and the streamer in both scenarios: a fixed IP network to be used as a reference for comparison purposes and a 5G virtualised infrastructure incorporating both VXLAN tunnelling for tenant isolation and GTP for

mobile phone mobility. PCAP files were gathered using TShark [43]. After the intensive validation tests, further extensive performance evaluation tests has been conducted with the testbed setting outlined in Figure 7 in order to test the scalability of the 5G-UHD FCA component and evaluate the impact of the virtualised 5G infrastructure on performance.

5.1   Performance Evaluation Testbed

The main purpose of this testbed is to measure stress behaviour and scalability of the 5G-UHD FCA component. The testbed is composed by three physical machines. The Streamer PC is the machine is charge of streaming simultaneously N times of the PCAP file used for the testbed. This PCAP contains the UHD reference video HoneyBee encoded in scalable 2K and 4K spatial resolutions using QP 27 and is being sent to the Client PC passing through the Network PC. Two different PCAP files have been used: IP PCAP file (1.4 MBytes) using an IP network with no encapsulation and 5G PCAP file (1.5 MBytes) using VXLAN over GTP encapsulation gathered from the operational testbed. The testbed ranges the number of simultaneous videos being transmitted from only one video flow to 512 simultaneous video flows in exponential increments of $N = (N - 1)^2$, i.e., (1, 2, 4, ..., 512). It is noted that 512 video flows add up to 1029 Mbps for 5 seconds, using an average bit rate for this theoretical estimation of the video flows at variable bit rates, which is the theoretical maximum achieved by the NIC interface. A higher number of video flows would cause saturation of the NIC and packets being drop in the streamer PC. It is worth remarking that all the experiments reported in this paper yielded an average of only 0.00017% of packet loss with a standard deviation of 0.0000002%. Therefore, packet loss is negligible.

The Streamer PC is an Intel Core 2 Duo CPU E8400 @ 3.00GHz, and is equipped with 8GB RAM, 500GB HDD SATA II, and 2x1 GBE NIC. The Network PC is an Intel Xeon CPU E5504 @ 2.00GHz, and has 24GB RAM, 500GB HDD SATA II with a clean Ubuntu 14.04 with forwarding capabilities enabled in the Linux kernel and the FCA software installed. Video flows were routed from the Network PC to the Client PC, which is composed by an Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GH, and has 2GB RAM, 500GB HDD SATA II with a clean Ubuntu 14.04 with the MP4Client tool of the GPAC software. The Client PC has also configured iptables to perform the mirroring of all the traffic being received from eth0 to eth1. This allows receiving in the Streamer PC all the traffic that the Client PC has processed/decoded. Using this setup, time measurement of both sent traffic and received traffic was performed in the same computer so that time synchronization is not an issue. Then, the PCAP files of both interfaces of the Streamer PC were stored and later on analysed in order to evaluate the performance of the 5G-UHD FCA.

Four different scenarios were carried out in this testbed. The first one, namely "IP FCA OFF", is the execution of all the scenarios without the FCA enabled using traditional IP traffic in order to create a reference. The second one "IP FCA ON" has the FCA enabled using traditional IP traffic. As the FCA has been configured to leverage Linux Netfilter, when it is switched on, iptables rules are installed in the kernel of the Network PC to adapt the Scalable H.265 video flows by selectively dropping the enhancement layer packets. Thirdly, the execution of all the scenarios without the FCA enabled using the 5G PCAP traffic was performed in order to create another baseline referred to as "VXLAN/GTP FCA OFF". Finally, the FCA was enabled using the 5G PCAP, referred to as "VXLAN/GTP FCA ON", in order to evaluate how the FCA behaves in a virtualised 5G network.
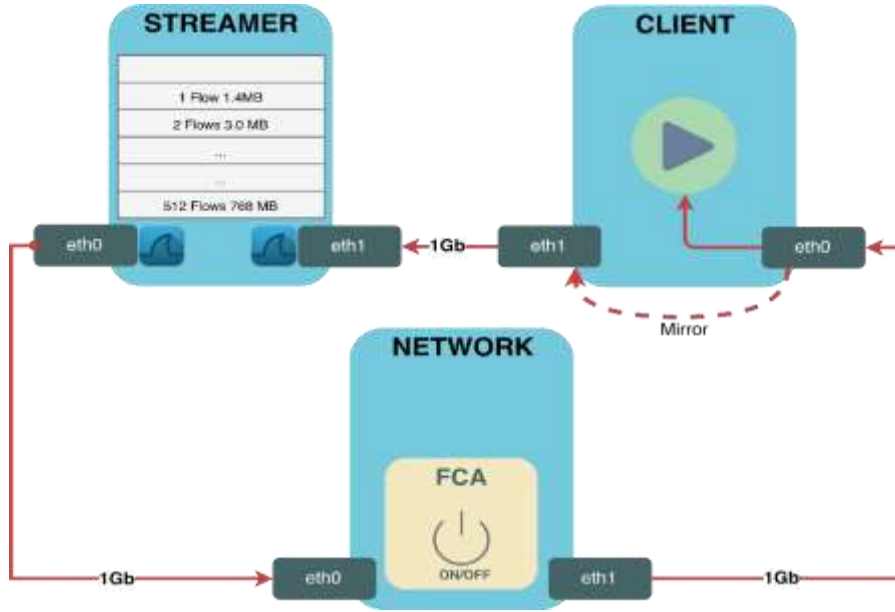
Figure 7. Testbed setting for performance evaluation. One to 512 5G video flows are streamed over this testbed to validate the operation performance and scalability of the proposed FCA in 5G and IP networks respectively.

5.2 Performance and Scalability Tests

Figure 8 shows the statistics of traffic delivery (sent and received), selective dropping and loss in the four scenarios executed. When the 5G-UHD FCA was disabled (see scenarios 1 and 3 in Figure 8), the total MBytes sent and those received were the same for all the experiments executed. In contrast, when the FCA was enabled (see scenarios 2 and 4 in Figure 8) and performed the video adaption by dropping the enhanced layer, in both IP and 5G networks the received traffic was 34% of the sent traffic leading to a bandwidth saving achieved by the FCA of 66% in average for all the experiments. These results have proved the efficiency of the FCA component. As highlighted before, such massive bandwidth saving was achieved without sacrificing the perceived quality (QoE) for end viewers.

With respect to the analysis of delay and jitter, henceforth expressed as delay ± jitter, the normal delivery of the video flows kept constant and exhibited close to 1 ± 0.1 ms performance when the FCA was disabled as shown in Figure 9. It is noted that both the IP and 5G networks showed very similar performance results, which indicates the high scalability of the 5G network being used as testbed. When the FCA was enabled, iptables rules were installed in the kernel and both delay and jitter were being increased in linear proportion to the number of rules (one rule per flow) with the increase of the number of flows. This is an expected behaviour. However, it must be noticed that the tests inserted not only traditional iptables rules for IP traffic but also the novel iptables rules for the virtualised 5G network traffic. In both cases, the results showed similar performance, which validates the suitability of the FCA for processing 5G network traffic, achieving similar performance comparable to that achieved in traditional IP networks despite the added overhead from additional 5G encapsulation showed in Figure 3. It is worth highlighting that even in the most loaded case in the system, i.e., when 512 simultaneous flows were being streamed, the system produced an average end-to-end delay of 25 ± 4.5 ms for the IP network whilst 25 ± 3.5 ms for the 5G network.
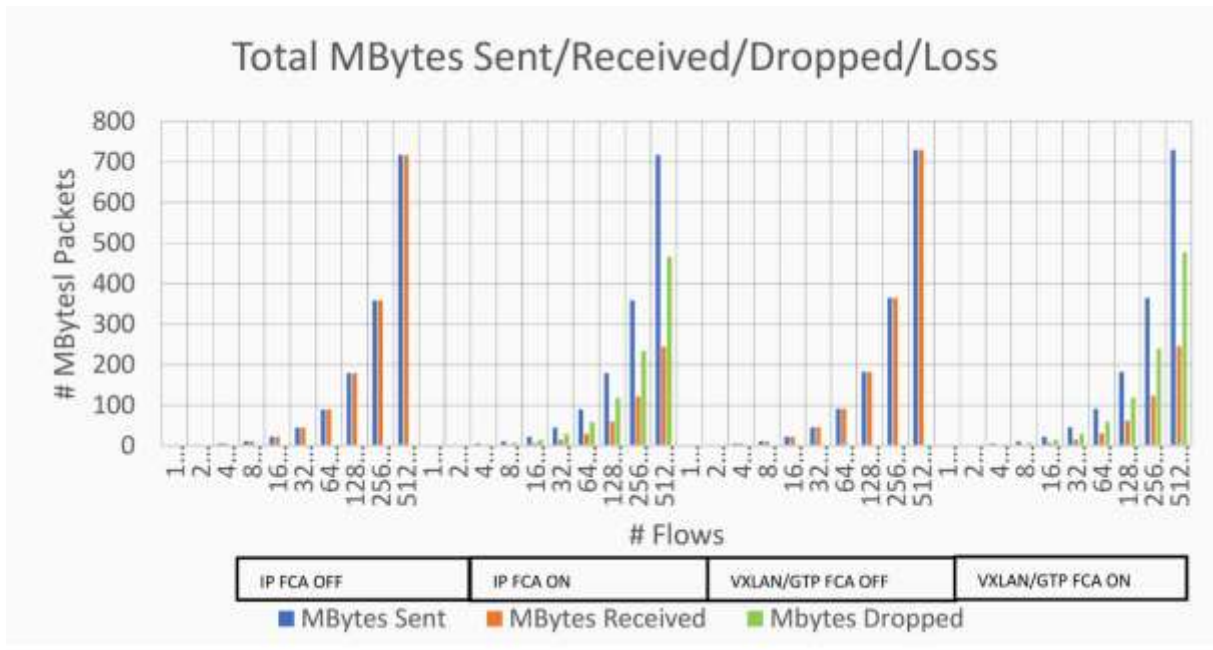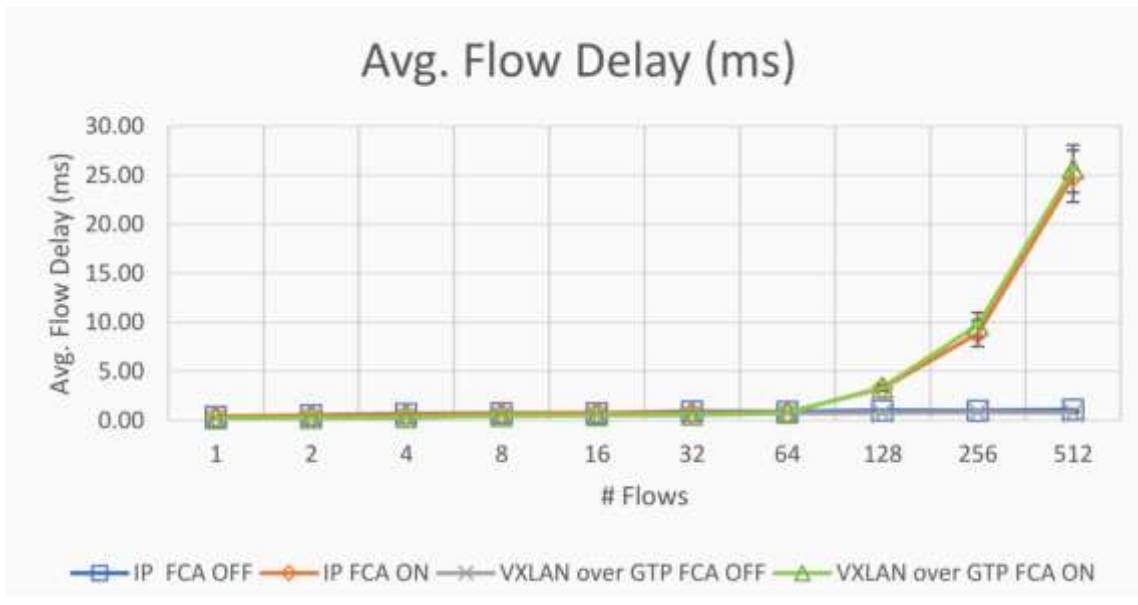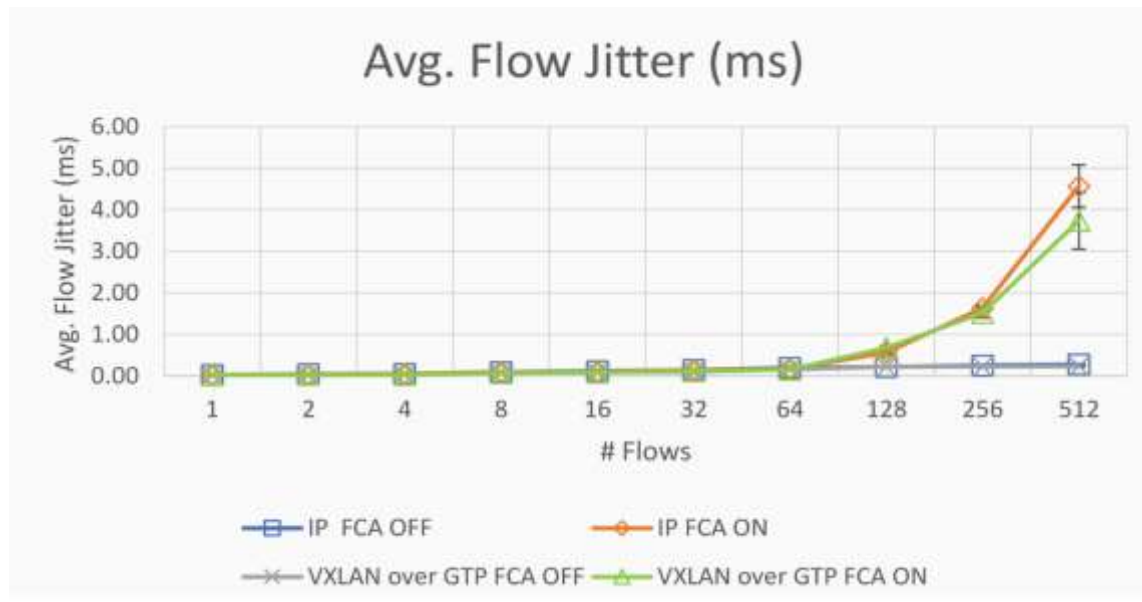
Figure 8 Total traffic sent/received/dropped/loss in four different scenarios: FCA switched on or off over pure IP networks and LTE-based 5G networks respectively.

Figure 10 compares directly and in more details the delay and jitter performance of the FCA adaptation of Scalable H.265 flows in the 5G network and the IP network. The negative values in Figure 10 mean that the delay or jitter was reduced when performing the FCA adaption in the 5G network with respect to the IP network. It is interesting to observe that both 5G scenarios, i.e. FCA enabled and disabled, actually yielded a lower average delay than that in the IP network when the number of flows was on the lower end. This finding is not intuitive. A further investigation reveals that the main reason for this performance improvement in 5G is related to the size of the packets. 5G networks use double encapsulation VXLAN/GTP, totalling 86 bytes extra overhead compared with the traditional IP traffic. This overhead added to every packet in this case favourably increases the inter-packet time between packet arrivals, leading to more time for the host machine to process the kernel interruptions incurred by packet arrivals and thus reducing processing delay and jitter in the host machine. On the other hand, when the system was being stressed (from 128 simultaneous flows onwards), the overhead delay becomes positive yet is still always below 1ms. This millisecond of delay is a very acceptable overhead in the worst-case scenario incurred by the 5G-UHD FCA component working in the 5G network.

(a)

(b)

Figure 9. Average flow delay (a) and flow jitter (b) in four different scenarios: FCA switched on or off over pure IP networks and LTE-based 5G networks respectively.

The jitter results in Figure 10 show that the jitter was kept almost the same in both IP and 5G networks and also in both scenarios with 5G FCA enabled or disabled. The results indicate a high scalability in the jitter performance even under stressed conditions of the proposed architecture. It is also noted that when the 5G FCA was enabled and the system was stressed (i.e., 256 and 512 simultaneous flows), it shows a lower average jitter in 5G network than in IP networks. The reason is the same as explained previously in the analysis of the delay.
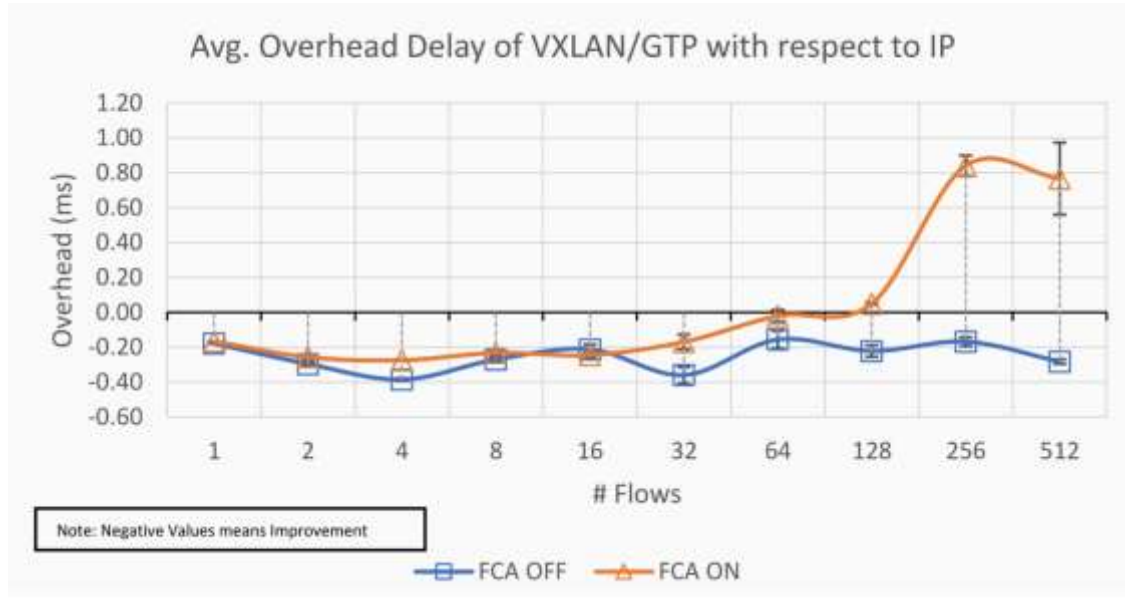
(a)
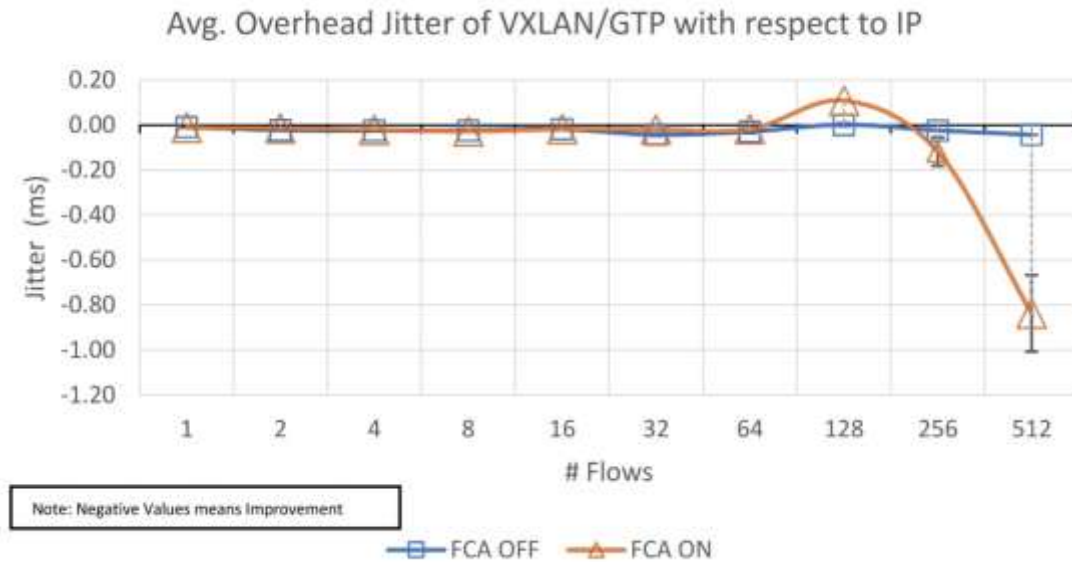
(b)

Figure 10. Comparison of average overhead delay (a) and jitter (b) of the video flows over 5G (VXLAN/GTP) networks with respect to over pure IP networks.

5.3   *Comparative Analysis*

To highlight the contributions of this paper, we have conducted a comparative analysis with a number of existing related studies. A range of technical aspects have been employed to compare our proposed system 5G-UHD with the relevant work, as shown in Table 1. These technical aspects cover video and streaming specific factors such as spatial resolution, codec, QoE awareness, video adaptation approach, streaming protocols etc., 5G networking technologies such as Cloud-RAN, multi-tenancy support, 5G encapsulation (GTP) support etc., and implementation (realistic testbed, emulation or simulation).

Based on this analysis as shown in Table 1, we can conclude that the proposed 5G-UHD approach and implementation offers a very promising, practical and more complete solution to achieve next-generation codec (Scalable H.265) based, QoE-aware UHD video streaming adaptation with 5G compliant capabilities, being fully operational in a realistic 5G networking testbed that integrates Cloud-RAN with a real RRH, LTE-based core network, multi-tenancy etc., in contrast with the existing work in this field.

Table 1 Comparative analysis of HD/UHD video streaming over 5G networks. This table compares the 5G-UHD in this work with representative related studies in the literature in terms of a range of technical design, implementation and evaluation aspects.

	Spatial Resolution (HD/FHD/UHD)	Codec	Video Adaptation Tech.	Video Adaptation Tech. Implementation	Streaming protocol	QoE-informed design?	User Equipment	RAN Tech.	RAN Implementation Setup	Cloud-RAN	Cloud RAN Implementation Setup	5G Edge-Core Segment	5G Edge-Core Implementation Setup	LTE/5G GTP Encapsulation Aware Adaptation	5G Cloud Multi-Tenant-Isolation
[54]	1080p (FHD)	NS	Filter module in medium access control	Adaptive bitrate	MPEG-DASH	Yes, Objective	NS	LTE	Emulation (OAI)	NC	NC	Based on MEC concept	Emulation	NC	NC
[55]	(UHD)	NS	Mobile edge virtualization	Adaptive prefetching	MPEG-DASH	Yes, Objective	Huawei Nexus 6P	LTE-A	Emulation	Yes	Yes	Based on MEC concept	Yes	NC	NC
[56]	NS	NS	Variable bit rate transcoding	Based on FFMPEG	RTP based	Yes, Objective	Laptop with 4G LTE USB Adapter, which includes a USIM card	LTE	Emulation (OAI)	Yes	NC	LTE EPC (HSS,MME,S GW)	Emulation	Yes, LTE (GTP)	NC
[57]	(UHD)	Scalable H.265	NC	NC	MPEG-TS based	NC	4K TV	NC	NC	NC	NC	NC	NC	NC	NC
[58]	NS	Scalable H.264	Expanding or dismissing the total number of SVC layers in each flow bit stream	Framework called TLS-AV	NS	Yes, Objective	NS	LTE	Simulation (LTE-Sim)	NC	NC	NS	Simulation	NC	NC
[59]	720p (HD)	H.264	Optimal bit rate encoder	NS	RTP based	Yes, Objective	NS	LTE	Simulation (LTE-Sim)	NC	NC	LTE EPC (HSS,MME,S GW)	Simulation	No, (LTE-Sim) Random direction model used instead	NC

[60]	720p (HD)	H.265	cross-layer scheduling framework	NS	RTP based	Yes	NS	LTE-A	Simulation	NC	NC	LTE-A EPC (HSS,MME,SGW)	Simulation	NC	NC
[61]	4K (UHD)	Scalable H.265	media-aware selective packet processing	NS	NS	Yes	NC	LTE-A	Theoretical	NC	NC	NC	NC	NC	NC
[62]	720p (HD)	Scalable H.265	Picture prioritization method and error concealment mode signaling	Custom encoder and signaling method to indicate the best Error concealment to apply.	RTP based	Yes, Objective	NC	NC	NC	NC	NC	NC	NC	NC	NC
[63]	576p (4CIF)	Scalable H.264	Media Aware Network Element	NS	RTP based	Yes, objective and subjective	1920x1080 screen	LTE	Emulation (OpenEPC)	NC	NC	NC	NC	NC	NC
[64]	768p (XGA)	Multiview H.265	Media Aware eNodeB packet prioritization	Medium Access Control (MAC) layer scheduler based	NS	Yes, Objective	NC	LTE	Simulation (IMTAphy)	NC	NC	NS	NS	NC	NC
[65]	NS	Scalable H.264	Low latency, low loss streaming mechanism using in-network coding and caching named L4C2.	By extending CCN (Content Centric Networking) approach	NS	Yes, Objective	NC	NC	Simulation (ndnSim)	NC	NC	NC	NC	NC	NC
5G-UHD (This paper)	4K (UHD)	Scalable H.265	media-aware selective packet processing	NetFilter based	RTP based	Yes, Objective and Subjective.	4K 5.5' mobile phone + 4K 55' TV	LTE (2x2 MIMO)	Emulation	yes	Emulation	LTE EPC (HSS,MME,SGW)	Emulation	Yes, LTE (GTP)	Yes (VXLAN)

NC - Not Considered
NS - Not Specified

6. Conclusion

UHD video applications are a main driver for 5G networks, and it is high time to investigate optimised delivery of UHD videos over 5G networks. This paper focuses on UHD video adaptation in congested 5G areas such as hotspots. A novel 5G-UHD framework has been designed, prototyped and empirically validated and evaluated. The empirical results have shown that the central 5G-UHD video flow processing component FCA is capable of adapting the video flows without sacrificing the viewers' perceived quality whilst mitigating the congestion. Furthermore, the video adaptation operation of a 5G-UHD FCA does not introduce any significant negative impact on the system's performance in terms of delay, jitter and packet loss, even in a virtualised 5G network testbed. In fact, when the system is not stressed, better performance compared with that in a pure IP network can be achieved.

Future work will investigate the porting of the 5G-UHD framework in an HTTP/DASH based paradigm, and evaluate the performance and the effect of the screen size of user equipment on QoE. Advanced 5G networking testbeds with higher speed will also be looked into to validate and evaluate high scalability performance. In addition, machine learning and other intelligence technologies will be explored to improve the self-optimisation capabilities of this framework.

Reference

[1]   Global 5G actions, 2017, [online] Available: https://5g-ppp.eu/global-5g-actions/.

[2]   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.

[3]   5G PPP, "Advanced 5G Network Infrastructure for the Future Internet," 2013. [Online]. Available at https://5g-ppp.eu/wp-content/uploads/2014/02/Advanced-5GNetwork-Infrastructure-PPP-in-H2020_Final_November-2013.pdf.

[4]   5G PPP, "5G-PPP Key Performance Indicators (KPIs)," [Online]. Available at https://5g-ppp.eu/kpis/.

[5]   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

[6]   5G PPP Architecture Working Group, "5G PPP View on 5G Architecture", June/July 2016. [Online]. Available at https://5g-ppp.eu/white-papers/.

[7]   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/.

[8]   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)

[9]   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/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

[10] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard", IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[11] 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.

[12] 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.

[13] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, A. Vetro, "Standardized extensions of High Efficiency Video Coding (HEVC)", IEEE J. Sel. Topics Signal Process., vol. 7, no. 6, pp. 1001-1016, Dec. 2013.

[14] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramonian, "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] 5G PPP, "Innovations for new Business Opportunities", 2017, [online] Available: https://5g-ppp.eu/wp-content/uploads/2017/01/5GPPP-brochure-MWC17.pdf.

[16] A. Gupta and R. K. Jha, "A Survey of 5G Network: Architecture and Emerging Technologies," *IEEE Access*, vol. 3, pp. 1206-1232, 2015.

[17] T. Schierl, M. M. Hannuksela, Y. K. Wang and S. Wenger, "System Layer Integration of High Efficiency Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1871-1884, Dec. 2012.

[18] Information Technology—Dynamic Adaptive Streaming Over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats, ISO/IEC 23009-1:2012, 2012.

[19] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62-67, April 2011.

[20] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469-492, Firstquarter 2015.

[21] IETF RFC 7798, "RTP Payload Format for High Efficiency Video Coding (HEVC)", 2016, [online] Available: https://www.ietf.org/rfc/rfc7798.txt.

[22] 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.

[23] M. Xing, S. Xiang and L. Cai, "A Real-Time Adaptive Algorithm for Video Streaming over Multiple Wireless Access Networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 795-805, April 2014.

[24] R. Parois, W. Hamidouche, E. G. Mora, M. Raulet and O. Deforges, "Real-time UHD scalable multi-layer HEVC encoder architecture," *2016 24th European Signal Processing Conference (EUSIPCO)*, Budapest, 2016, pp. 1298-1302.

[25] W. Hamidouche, M. Raulet and O. Déforges, "4K Real-Time and Parallel Software Video Decoder for Multilayer HEVC Extensions," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 169-180, Jan. 2016.

[26] "Sprint, Ericsson Demo 4K UHD Video Streaming on Live 5G Wireless", 2016, [online]

Available: https://www.thefastmode.com/technology-solutions/8349-sprint-ericsson-demo-4k-uhd-video-streaming-on-live-5g-wireless.

[27] "DoCoMo and Nokia made first live 8k UHD video transmission via 5G", 2016, [online] Available: http://nextvnews.com/docomo-nokia-made-first-live-8k-uhd-video-transmission-via-5g/.

[28] UHD-on-5G project, 2016, [online] Available: https://team.inria.fr/diana/uhd-on-5g/.

[29] V. Poliakov, L. Sassatelli and D. Saucez, "Case for Caching and Model Predictive Control Quality Decision Algorithm for HTTP Adaptive Streaming: Is Cache-Awareness Actually Needed?," *2016 IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, 2016, pp. 1-6.

[30] 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.

[31] T. Biatek, W. Hamidouche, J. F. Travers and O. Deforges, "Adaptive rate control algorithm for SHVC: Application to HD/UHD," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 1382-1386.

[32] X. Huang, M. M. Hannuksela and H. Li, "Improved downstream rate-distortion performance of SHVC in DASH using sub-layer-selective interlayer prediction," *2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*, Xiamen, 2015, pp. 1-6.

[33] W. Hamidouche, G. Cocherel, J. Le Feuvre, M. Raulet, and O. Deforges, "4K real time video streaming with SHVC decoder and GPAC player", Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on, Jun 2014, Chengdu, China. pp.1-2, 2014.

[34] OpenAirInterface, 2017, [online] Available: http://www.openairinterface.org/.

[35] OpenStack, 2017. [Online]. Available at https://www.openstack.org/.

[36] IETF RFC 7348, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", 2014, [Online]. Available: https://tools.ietf.org/html/rfc7348.

[37] The Netfilter project, 2017, [Online]. Available: https://www.netfilter.org/.

[38] SHVC reference software, 2017, [Online]. Available: https://hevc.hhi.fraunhofer.de/shvc.

[39] GPAC, 2017, [Online]. Available: https://gpac.wp.imt.fr/.

[40] ETSI, Multi-access Edge Computing (MEC), [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing.

[41] 4K video sequences, 2017, [Online]. Available: http://ultravideo.cs.tut.fi/#testsequences.

[42] OpenDayLight, 2017, [Online]. Available: https://www.opendaylight.org/.

[43] TShark, 2017, [Online]. Available: https://www.wireshark.org/docs/wsug_html_chunked/AppToolstshark.html.

[44] RabbitMQ, 2017, [Online]. Available: https://www.rabbitmq.com/.

[45] Kafka, 2017, [Online]. Available: https://kafka.apache.org/.

[46] R. Sjoberg et al., "Overview of HEVC High-Level Syntax and Reference Picture Management," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1858-1870, Dec. 2012.

[47] 3GPP, "Evaluation of High Efficiency Video Coding (HEVC) for 3GPP services (Release

12)", TR 26.906 V12.0.0, 2014.

[48] DPDK, 2017, [Online]. Available: http://dpdk.org/.

[49] NetFPGA, 2017, [Online]. Available: http://netfpga.org/.

[50] Wireshark, 2017, [Online]. Available: https://www.wireshark.org/.

[51] Recommendation ITU-R BT.500-13, Methodology for the subjective assessment of the quality of television pictures, Jan 2012.

[52] JSON, 2017, [Online]. Available: http://www.json.org/.

[53] AMQP, 2017, [Online]. Available: https://www.amqp.org/.

[54] C. C. Wang, Z. N. Lin, S. R. Yang and P. Lin, "Mobile edge computing-enabled channel-aware video streaming for 4G LTE," 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, 2017, pp. 564-569.

[55] C. Ge, N. Wang, G. Foster and M. Wilson, "Toward QoE-Assured 4K Video-on-Demand Delivery Through Mobile Edge Virtualization With Adaptive Prefetching," in IEEE Transactions on Multimedia, vol. 19, no. 10, pp. 2222-2237, Oct. 2017.

[56] M. A. Kourtis, H. Koumaras, G. Xilouris and F. Liberal, "An NFV-based Video Quality Assessment Method over 5G Small Cell Networks," in IEEE MultiMedia, vol. PP, no. 99, pp. 1-1.

[57] R. Parois, W. Hamidouche, E. G. Mora, M. Raulet and O. Deforges, "Demo: UHD live video streaming with a real-time scalable HEVC encoder," 2016 Conference on Design and Architectures for Signal and Image Processing (DASIP), Rennes, 2016, pp. 235-236.

[58] S. Kumar, S. Sriram, A. Sarkar and A. Sur, "A three level adaptive video streaming framework over LTE," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, pp. 002442-002447.

[59] H. Fu, H. Yuan, M. Li, Z. Sun and F. Li, "Models and analysis of video streaming end-to-end distortion over LTE network," 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 516-521.

[60] A. Jassal and C. Leung, "H.265 video capacity over beyond-4G networks," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6.

[61] O. Awobuluyi, J. Nightingale, Q. Wang and J. M. Alcaraz-Calero, "Video Quality in 5G Networks: Context-Aware QoE Management in the SDN Control Plane," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, 2015, pp. 1657-1662.

[62] E. S. Ryu, Y. Ryu, H. J. Roh, J. Kim and B. G. Lee, "Towards robust UHD video streaming systems using scalable high efficiency video coding," 2015 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2015, pp. 1356-1361.

[63] C. Mysirlidis, A. Lykoyrgiotis, T. Dagiuklas, I. Politis and S. Kotsopoulos, "Media-aware proxy: Application layer filtering and L3 mobility for media streaming optimization," 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 6912-6917.

[64] A. Jassal, B. Oztas, M. T. Pourazad and P. Nasiopoulos, "A packet prioritization scheme for 3D-HEVC content transmission over LTE networks," 2015 IEEE International Conference on Communication Workshop (ICCW), London, 2015, pp. 1788-1793.

[65] Kazuhisa Matsuzono, Hitoshi Asaeda, Thierry Turletti. Low Latency Low Loss

Streaming using In-Network Coding and Caching. IEEE INFOCOM, May 2017, Atlanta, United States.