Enabling Agile Video Transcoding over SDN/NFV-enabled Networks

Harilaos Koumaras, Christos Sakkas, Michail Alexandros Kourtis, Christos Xilouris, Vaios Koumaras
Institute of Informatics and Telecommunications
NCSR "Demokritos"
Athens, Greece

Georgios Gardikis
R&D Department
Space Hellas S.A.
Athens, Greece

***Abstract*— Software Defined Networking (SDN) and Network Function Virtualization (NFV) provide an appealing 5G vision of how content distribution networks can be enhanced with in-service transcoding and adaptation processes. This paper describes a proof-of-concept experimental implementation of SDN/NFV-enabled network domain towards providing an agile video transcoding process for maintaining the QoE level of a media service when network congestion occurs. The paper shows how SDN/NFV techniques are facilitating the dynamic deployment of a transcoder VNF upon a triggering event of a congested bottleneck at the access link of the end-user. Video quality measurements demonstrate the efficiency of the VNF deployment and the seamless applicability of the transcoding process without experiencing any service interruption.**

Keywords—NFV; SDN; 5G; multimedia; QoE; video quality.

I. Introduction

Currently network operators are gradually transiting towards the 5G networks, which are envisaged to consist of heterogeneous wireless and wired physical infrastructures, whose resources are unified and dynamically pooled and offered in as-a-Service fashion to multiple tenants [1]. In this evolution, the operators have to deal with the continuously increasing demand for real-time entertainment services (e.g. streaming video and audio), which at the moment counts for at least the 60% of the total Internet traffic [2]. Considering the expected network bandwidth increase in 5G networks [3] in conjunction with the emergence of the ultra-high definition videos, it is evident that the consumer demand in the next years for even higher quality media content will increase rapidly. Adding to this trend the impetus by the development of novel value-added services within the 5G framework, then it becomes evident that together with the rapid growth of real-time entertainment services, novel techniques for self-adaptation and self-optimization of the provided video streams should be designed, developed and deployed.

Towards pursuing these self-optimization techniques, a number of new technological advances, which are changing the conventional way of provisioning and operating the networks, should be taken under consideration as enabling technologies.

Cloud-based self-optimization techniques has been proposed [4][5], which receive the multimedia content from a live source and transcode it in real-time (actually much faster than real-time is possible) into various content representations based on a given configuration (e.g., video profile, video quality, video resolution, audio/video bitrate). Although, the main advantage of the cloud-based optimization techniques is the resource elasticity that provides high performance of the process when needed, however the main disadvantage relies to the fact that the topology of the cloud-based transcoder is static within the network, without allowing flexible and on-demand deployment of the self-optimization technique when needed and close to the user at the edge of the network.

Therefore, in order to advance further the elasticity of the cloud-based self-optimization techniques with more flexibility and agility in the adaptation technique provision and instantiation, two of the major trends in the telecom industry today are considered: The Software Defined Networking (SDN) and the Network Functions Virtualization (NFV).

SDN proposes to decouple the control and data planes in network nodes, logically centralizing that control, while leaving the network elements to forward traffic and enforcing the control policies according to instructions received from the centralized controller. This makes the network programmable in a way that promises to be more flexible than the current managed paradigm.

On the other hand, NFV envisages the instantiation of network functions on commodity hardware, breaking the monolithic approach to functional software and hardware that exists in today's vendor offerings. Although they are separate initiatives it appears that SDN and NFV are complimentary and are considered as a key enabling technology for the provision of value added QoE-driven services [6] in the forthcoming 5G networks, such as high-bitrate and high-quality 2D/3D broadcast content, coupled with interactive personalized services [7].

In this framework, this paper presents a proof-of-concept SDN/NFV-enabled experimental testbed, which implements a self-optimization and self-adaptation Virtual Network Function (VNF) of real-time video streams, adapting on the fly the encoding characteristics of the provisioned video stream. The objective of the paper is to present and experimentally validate how the SDN and NFV can enable the provision of novel in-service adaptation and transcoding techniques.

The rest of the paper is organized as follows: Section II describes the SDN and NFV as enabling technologies for video adaptation actions. Section III describes implementation aspects and challenges of the architecture in terms NFV managements and orchestration. Section IV describes the proof-of-concept experimental testbed and the experimental results of the proposed SDN/NFV-based agile transcoding process. Finally Section V concludes the paper.

II. SDN AND NFV AS ENABLING TECHNOLOGIES

In this paper, we examine how SDN and NFV can be used in a complimentary way as enabling technologies for real-time in-service video transcoding process. This section provides a brief overview of the SDN/NFV technologies.

The huge growth of virtualisation and cloud computing in the market prompts the IT industry to revise their opinions about current traditional network architectures. The majority of traditional networks are built with Ethernet switches, which are arranged in a tree structure. This design was adequate in the time that client-server computing was at the forefront. But currently things have changed and the concept of virtualisation has entered at various layers of the computing industry, spanning from the application/service layer down to networking and function virtualisation. NFV appears as an emerging aspect in the networking domain, which has the potential to radically redefine the substance of what is referred to as “network infrastructure”. NFV refers to the virtualization of network functions, which are called Virtual Network Functions (VNFs), carried out today by specialized hardware devices and their migration and deployment as software-based appliances (i.e. VNFs) on top of commodity IT infrastructures, such as cloud computing infrastructures. NFV offers certain obvious advantages, such as rapid deployment of the network functions, easy relocation and resource upgrading by reducing the reliance on proprietary devices and improving service flexibility by using a more agile software-based framework for building service features.

This virtualised heterogeneity creates new needs for management and administration, capable of handling the new virtualised functions and optimizing the respective resource availability. This new technology is called Software Defined Networking (SDN), which enables dynamic and flexible traffic steering and forwarding, leads to higher application performance and business efficiency. Although these operations can be achieved via traditional networking technologies, their implementation via SDN offers much more flexibility and agility. Among the advances of applying flexible traffic steering and switching, supported by an SDN architecture with a programmatic northbound API are the centralized and granular control over the federated networking infrastructure, enabling customers to experience dynamic changes to the access network through which the requested service is delivered, either immediately depending on specific triggering conditions or scheduled in the future.

In the proposed architecture NFV implements network functions that are typically deployed in specialized hardware servers as software instances running on commodity servers through software virtualization techniques. These virtualized functions are called VNFs, and can be located in the most appropriate places (e.g., data centers), referred to as NFV Infrastructure Point of Presence (NFVI-PoPs). NFV is applicable to any data plane packet processing and control plane function in fixed and mobile network infrastructures.

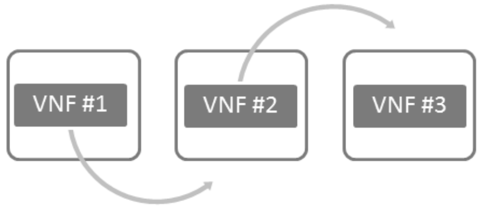

Fig. 1. Concept of Service Function Chaining utilizing SDN techniques

NFV is complementary to SDN: while network functions can be virtualized without the need of an underlying SDN infrastructure, both are mutually beneficial towards Service Function Chaining (SFC), which is a technique for selecting and steering data traffic flows with SDN rules through various VNFs as Fig.1 depicts.

III. IMPLEMENTATION ASPECTS AND CHALLENGES

In order to derive a better insight of the hierarchical logic of the proposed SDN/NFV-enabled network, the architectural figure can be transposed from a domain-centric to a layered view.

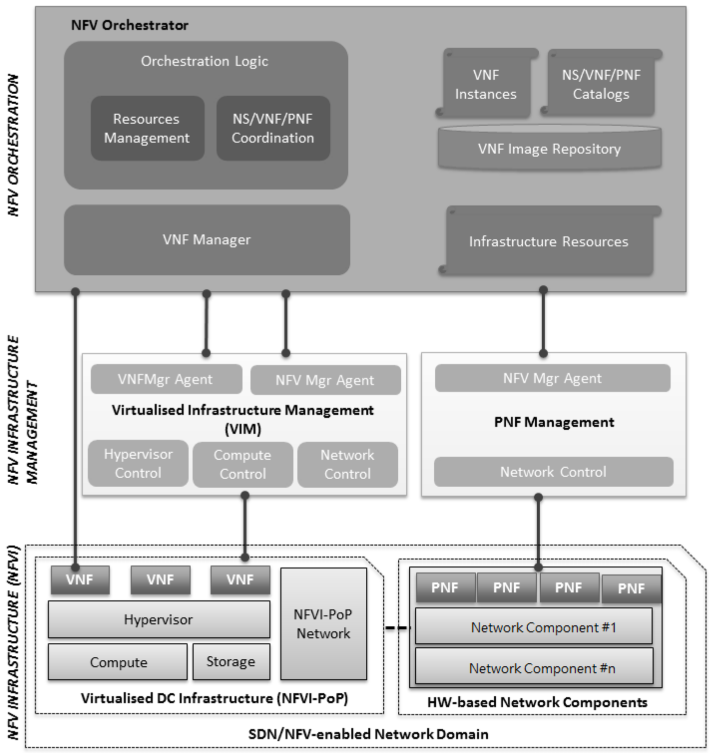

Fig. 2. SDN/NFV-enabled network domain layered approach

Fig. 2 shows that the SDN/NFV-enabled network domain is split into three logical layers:

- The Infrastructure (lower) layer includes the virtualization-capable equipment on which the network service is deployed. This layer includes:
  - The SDN and non-SDN network elements of the terrestrial Wide Area Network (transport and edge)
  - The distributed NFVI-PoPs (data centres with compute clusters with the supporting SDN network).
  - The Physical Network Functions (PNFs), which are hardware-based network elements performing a specific networking function (e.g. HW-based router etc).
- The Infrastructure Management (middle) layer includes distributed management entities for the various parts of the infrastructure. The SDN/NFV enabled segments (NFVI-PoPs) are managed by a Virtualised Infrastructure Management (VIM) entity. In addition to the VIM, a legacy management entity (i.e. PNF Management) needs to be foreseen for the management of the non-SDN/NFV components; while it is beneficiary to employ SDN at several parts of the WAN network, support of legacy non-SDN elements is deemed necessary to ensure interoperability.
- The Orchestration (top) layer is responsible for the coordination of the entire administrative domain, the infrastructure as well as the services which run on it. It orchestrates virtualised resources to compose end-to-end services and optimizes them dynamically. This role is undertaken by an Orchestrator entity, which normally closely interacts with (or ideally is integrated in) the operator's overall Network Management System (NMS).

However, although SDN is a key enabler for network virtualization, partitioning and dynamic control, various risks and challenges are still pertinent. By means of a single or distributed SDN controllers, the network operator can manage the partitioning of the network and offer the slices to several Content Providers (CPs). On top of that, these slices can also be programmable; the CP can develop an arbitrary SDN application which will control the hybrid virtual network and manipulate/divert the media streams across multiple paths as desired. This is the so-called "SDN as-a-Service" (SDNaaS) service paradigm [8], which is a significant added-value compared to static, non-programmable virtualization, in which the CP just uses the offered capacity, without any control capabilities. However, in this case, SDN security aspects need to be taken into account, since the network control applications should neither affect the stability of the infrastructure resources, nor interfere with (possible) other CPs using the same infrastructure.

The use of SDN brings benefits not only to network control, but also to network monitoring. Sets of network metrics, such as per-flow latency, loss etc. can be provided to the CP in real time, so that the latter can dynamically decide the balancing of the load between the available paths. In any case, the exposure of these metrics should be done in a controlled manner, so as not to affect the privacy of the network operator and to avoid exposure of sensitive data about the status of the infrastructure.

IV. Proof-Of-Concept Experimental Testbed

The experimental topology of the proposed architecture is depicted in Fig. 3, where at the ingress and egress points of the terrestrial network domain have been placed two SDN-compatible Open Virtual Switches (OVS), which are under the management and control of the OpenDaylight SDN controller. At the engress point of the domain, an Opestack cloud platform is installed to support a NFVI-PoP, which is capable upon appropriate orchestration to instantiate VNFs and apply SFC among them.

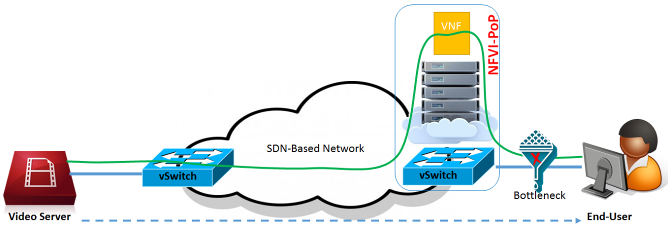

Fig. 3. Topology of the experimental testbed

The executed experiment will demonstrate SDN-based video service steering through the NFVI-PoP, showing the advantages of the NFV applicability in the real time video transcoding process. In the simplest approach, the CP just uses the SDN/NFV-enable network domain as a "dumb pipe" (yet with specific SLA) to convey media streams. However, a significant added-value of the use of virtualization and programmability technologies would be to offer to the CP elevated management and control capabilities on the SDN/NFV-enabled network domain. This means that the CP may develop his/her own network control logic in order to dynamically configure the network at runtime, allocate resources and also influence routing/forwarding decisions as desired.

In this framework, a video server is considered in the testbed, which hosts the videos at their original encoded form. In order to maintain an acceptable video service delivery, we consider that our testbed is equipped with a real time video adaptation system, which is capable of performing transcoding in real-time, which results in video streams adapted to the current network conditions, but of degraded video quality. This video transcoder is implemented for the needs of the paper based on the widely used FFMPEG and is instantiated as VNF at the egress NFVI-PoP. The adaptation triggering event may differ depending on the specific use case, ranging from the user terminal specifications to the available bandwidth of the delivery channel. For the scope of this paper, the triggering event is manually controlled in order to perform the experimental validation of the proposed NFV-based transcoder, based on congestion bottleneck that is artificially created at the access link of the end-user.

Therefore, for the experimental needs of the paper, an appropriate SDN-compatible software program was developed, which achieves appropriate traffic steering of the video flow through the NFVI-PoP when congestion bottleneck is monitored at the access link in order the video to be transcoded in real time and seamlessly by the transcoder-VNF, which has been instantiated there.

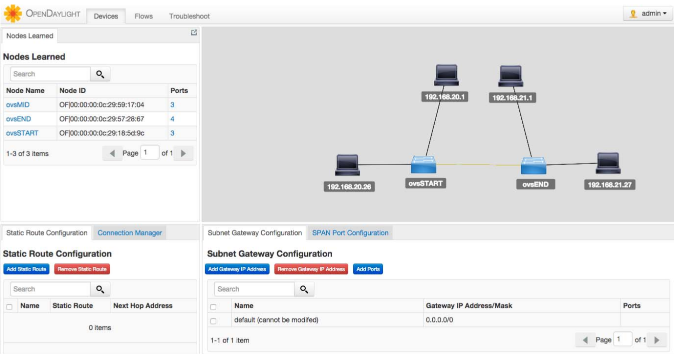

Fig. 4. The experimental topology at the SDN Controller GUI

Fig. 4 depicts the experimental topology of the testbed as it is demonstrated by OpenDaylight platform [9], which provides a unified management platform of SDN-enabled domain.

This demo scenario considers that one end user wishes to receive video content, therefore one unicast flow is initiated by the CP (i.e. the Video Server). However, the requested media stream over-pass the available bandwidth in the access link of the end-user, which has been on purpose lowered to facilitate the execution of the experiment, resulting in congestion bottleneck and therefore degradation of the QoE of the transmitted media signals.

Towards facilitating the video transmission in a dynamic and transparent way for the end-users, the orchestrator monitors the quality degradation and by following an edge computing strategy instantiates a transcoder as VNF at the NFVI-PoP and applies appropriate SDN-based traffic steering rules at the OVSs in order to support the Service Function Chaining (SFC), i.e. to steer the media flow through the VNF-based transcoder and then the transcoded at lower bit-rate signal to be forwarded at the end-user. The steered traffic is represented by the green line at Fig.3. The storyline of this scenario is the following:

- An end-user requests a unicast media service
- Total traffic (i.e. the requested media and background traffic of other services) exceeds the available bandwidth of the access link resulting to network congestion and quality degradation.
- Orchestrator monitors and instantiates at the NFVI-PoP located at the edge, a VNF-based transcoder and appropriate SDN rules for the traffic steering in order to support the SFC.
- The delivered media stream is transcoded (i.e. at lower bitrate/frame rate/resolution) in real time and transparently from the end-user.
- The transcoded media service can fit in the available bandwidth of the access link and is transmitted without causing any network congestion/bottleneck.
- The QoE level of the transmitted signals is re-instated at satisfactory level and the signal reach the end-user without impairments.

Across all the experiment, the QoE level is measured at the user side utilizing the SSIM metric [10]. Initially, a unicast video is initiated from the video server at the port 33334.

```
root@enduser:~# ffmpeg -re -i big_buck_bunny_720p_surround.avi -vcodec mpeg4 -an
 -b 512k -s 320x280 -f mpegts rtp:192.168.20.26:33334
ffmpeg version 0.8.16-4:0.8.16-0ubuntu0.12.04.1, Copyright (c) 2000-2014 the Lib
av developers
  built on Sep 16 2014 18:33:49 with gcc 4.6.3
```

Fig. 5. Unicast video flow initiation at the media server

The requested video is approximately 512 kbps, which together with the background traffic created by other services, exceeds the total capacity of the access link, resulting to severe quality degradation at the end-user side as Fig. 6 depicts.

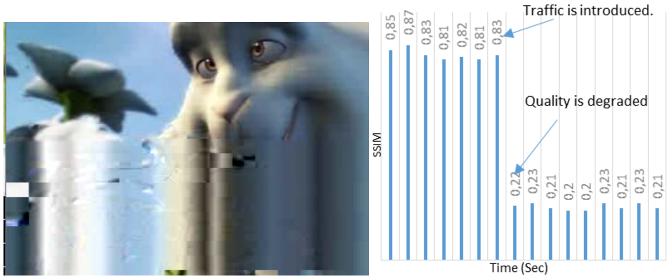

Fig. 6. Degraded media delivery and observed video quality at SSIM values

Towards dynamically and seamlessly improving the media service delivery, the orchestrator instantiates the transcoder VNF at the NFVI PoP, as it is depicted in Fig. 7, and performs transcoding in real time and seamlessly.

```
Input #0, mpegts, from 'udp:192.168.21.50:33334':
  Duration: N/A, start: 1.400000, bitrate: N/A
  Program 1
    Metadata:
      service_name    : Service01
      service_provider: Libav
    Stream #0.0[0x100]: Video: mpeg4 (Simple Profile), yuv420p, 320x280 [PAR 14:
9 DAR 16:9], 24 fps, 24 tbr, 90k tbn, 24 tbc
[buffer @ 0x1eab560] w:320 h:280 pixfmt:yuv420p
[mpegts @ 0x1efb440] muxrate VBR, pcr every 2 pkts, sdt every 200, pat/pmt every
 40 pkts
Output #0, mpegts, to 'udp:192.168.20.26:33334':
  Metadata:
    encoder         : Lavf53.21.1
    Stream #0.0: Video: mpeg4, yuv420p, 320x280 [PAR 14:9 DAR 16:9], q=2-31, 256
 kb/s, 90k tbn, 24 tbc
Stream mapping:
  Stream #0.0 -> #0.0
```

Fig. 7. VNF Instantiation and transcoding initiation of media service

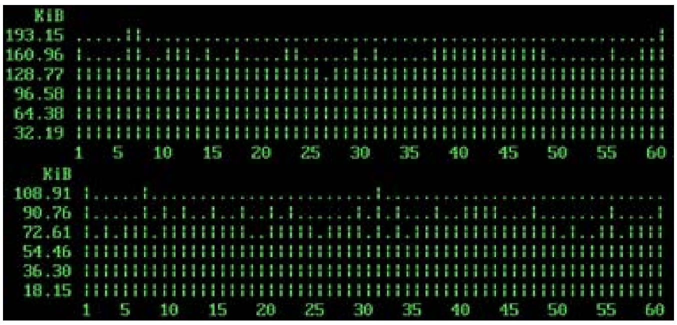

Fig. 8. Media Service Traffic Before and After Transcoding

It should be pointed out that for each media service a different instance of the transcoder VNF can be instantiated, which in case of multiple flows more than one instances should be executed at the NFVI-PoP. Fig. 8 depicts the total traffic before and after the transcoding process, as monitored at the Network Interface Controller (NIC) of the Virtual Machine (VM). Based on this stressful condition, Fig. 9 depicts an overall system load of 5% in terms of CPU utilization, 9.4% HDD utilization and 36% memory usage, allowing to the VM to operate on a stable and efficient status.

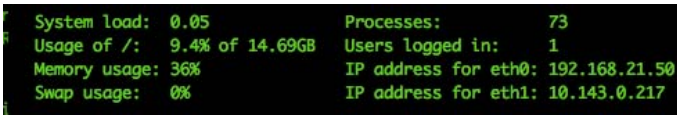

Fig. 9. Performance parameters (%CPU, %MEM, %HDD) of the VM hosting the VNF

Upon the real time transcoding of the media service from 512kbps down to 256 kbps, the video quality is reinstated seamlessly (i.e. without requiring any interruption) as Fig. 10 depicts, since now the total traffic is reduced and therefore is served efficiently through the available capacity of the access link.

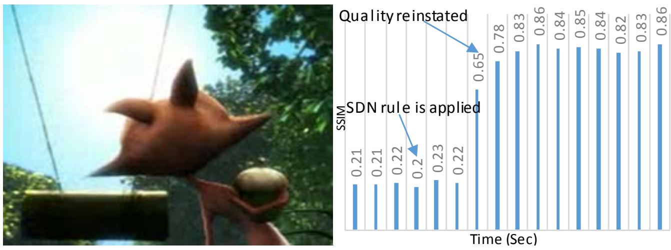

Fig. 10. Video Quality is reinstated upon SDN-based traffic steering over NFVI-PoP and SSIM values are increased accordingly

The responsiveness of the whole testbed after the triggering event due to the congestion is approximately 1.5 sec, during which the SDN-based traffic steering rules are applied and executed by the SDN switch and finally the QoE level of the delivered service is reinstated upon the VNF transcoding process.

V. Conclusions

This paper has examined the applicability of the NFV and SDN technology towards providing agile video adaptation solutions. Based on the presented proof-of-concept experimental testbed, it becomes evident that the support of the NFV and SDN paradigm seems crucial in order to maintain the competitiveness of the content provider within the forthcoming 5G ecosystem. The paper has shown that the agility offered via SDN/NFV-based infrastructure is a key factor to enhance video quality in case of network congestion with unprecedented flexibility and reconfigurability, as well as to develop novel adaptation services with additional added-value features.

Acknowledgment

This work has been partially funded from the European Space Agency ARTES 1 project CloudSat ("Scenarios for integration of Satellite Components in Future Networks" - ESA/ESTEC Contr. No.: 4000110995/14/NL/AD) and the European Union's VITAL project of the H2020 Research and Innovation Programme (H2020-ICT-2014-1) under the Grant Agreement H2020-ICT-644843.

References

[1] The Integral SatCom Initiative (ISI), "SatCom role in 5G networks", Net!Works Event 2013 Implementing H2020, Brussels, 29.10.13

[2] Sandvine, "Global Internet Phenomena Report 2H 2014", Sandvine Intelligent Broadband Networks, 2014.

[3] E. Liotou, H. Elshaer, R. Schatz, R. Irmer, M. Dohler, N. Passas, and L. Merakos, "Shaping QoE in the 5G ecosystem," in 7th International Workshop on Quality of Multimedia Experience (IEEE QoMEX), Costa Navarino, May 2015.

[4] Christian Timmerer, Daniel Weinberger, Martin Smole, Reinhard Grandl, Christopher Mueller, Stefan Lederer, "Live Transcoding and Streaming-as-a-Service with MPEG-DASH", in 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (Enrico Magli, Stefano Tubaro, Anthony Vetro, eds.), IEEE, Los Alamitos, CA, pp. 1-4, 2015.

[5] Christian Timmerer, Daniel Weinberger, Martin Smole, Reinhard Grandl, Christopher Mueller, Stefan Lederer, "Cloud-based Transcoding and Adaptive Video Streaming-as-a-Service", In IEEE Multimedia Communications Technical Committee E-Letter, IEEE Communications Society, New York, NY, USA, pp. 7-11, 2015.

[6] D. Tsolkas, E. Liotou, N. Passas, and L. Merakos, "The need for QoE-driven interference management in femtocell-overlaid cellular networks," Mobile and Ubiquitous Systems: Computing, Networking, and Services, Springer International Publishing, Editors: I. Stojmenovic, Z. Cheng, and S. Guo, ISBN: 978-3-319-11569-6, vol. 131, pp. 588-601, Sept. 2014

[7] M. Milosevic; B. Rodaljevic; M. Pejovic; M. Radonjic, "One implementation of companion screen functionality for hybrid broadcast broadband television," in Telecommunications Forum Telfor (TELFOR), 2015 23rd , vol., no., pp.783-786, 24-26 Nov. 2015

[8] B. Lee, S. H. Park, J. Shin, S. Yang, "IRIS: The Openflow-based Recursive SDN controller," in 16th International Conference on Advanced Communication Technology (ICACT), pp.1227-1231, 16-19 Feb. 2014

[9] OpenDaylight network controller, http://www.opendaylig

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity". IEEE Transactions on Image Processing, pp. 600–612, vol. 13 (4), April 2004