Orchestrating Live Immersive Media Services Over Cloud Native Edge Infrastructure


Abstract—Live streaming is becoming an essential for the entertainment and media industry. Quality of Service (QoS) and reliability is a major factor in live streaming, whereas availability and interaction between users are considered very important. Furthermore, the amount of data those services produce is huge, while the data production rate has dramatically increased over the last years. Thus, the use of 5G networking technologies, such as Network Function Virtualization (NFV), appears to be a promising solution for addressing the above-mentioned challenges. To this end, in this paper, we describe how an existing infrastructure of live streaming and social media services can take advantage of the NFV-enabled 5GTANGO Service Platform (SP). Also, in this paper we describe the capabilities of the SP and how it can support such services. To be more specific, an efficient monitoring framework, that supports the monitoring of Service Level Agreement (SLA) Management Framework. The aforementioned features incorporate mechanisms for collecting real-time monitoring data for a better management and optimization that guarantees the required Qos of the media service. The remaining of the paper is organized as follows. Section II presents the background and motivation of this work. Section III describes the immersive media service and its use cases. Section IV introduces the proposed 5GTANGO capabilities giving emphasis to the Monitoring and SLA Management frameworks for the specified immersive media use case respectively, while Section V discusses the evaluation results. Finally, in Section VI, we close up with some ideas on future extensions.

I. INTRODUCTION

There is a huge penetration of media in live events (e.g., stadiums), characterized by high-density areas that challenges the underlying 5G technologies to a different extent. During those large live events, the peaks of simultaneous viewing (e.g., sports events in stadiums) can be significant for a traditional network to support [1]. There has been huge research effort to remove the technical barriers surrounding the stream from individual cameras through 5G technologies, therefore providing better quality video with ability to introduce immersiveness. Sizable amount of data used in these events is in correspondence to the growing extent of social media. Users now require immersive experience but at the same time would like to be at the comfort of their home. The introduction of Augmented Reality/Virtual Reality (AR/VR) in the consumer market will help in providing this experience. Quality of Service (QoS) has to be a top priority in immersive media though, while availability and interaction between users are considered to be critical [2]. The immersive media is defined as an integrated representation of various media components that can offer participating users presence and immersion. Therefore, the immersive media use case implemented by the 5GTANGO Project, aims to provide a new adaptive and immersive end-to-end streaming solution which is able to fuse multi-source video stream from 360 and non-36 cameras. The main goal of the Immersive Media Network Service (NS) that is going to be described in this papers, is to provide a live experience to the audience, which is also enhanced by a personalized social media feed for each user by adopting the 5GTANGO capabilities, along with the advanced Monitoring Management Framework and an Service Level Agreement (SLA) Management Framework. The aforementioned features incorporate mechanisms for collecting real-time monitoring data for a better management and optimization that guarantees the required Qos of the media service. The remaining of the paper is organized as follows. Section II presents the background and motivation of this work. Section III describes the immersive media service and its use cases. Section IV introduce the proposed 5GTANGO capabilities giving emphasis to the Monitoring and SLA Management frameworks for the specified immersive media use case respectively, while Section V discusses the evaluation results. Finally, in Section VI, we close up with some ideas on future extensions.

II. BACKGROUND AND MOTIVATION

It is doubtless that active and advanced work on media services in 5G ecosystem is carried out during the last years [3]. Media applications pose specific network demands in order to deliver seamless operation to the end users and unlock the full potential of media applications in the 5G ecosystem [4]. It is worth mentioning that 5G enables events (e.g. football games) to be enhanced by new immersive experiences that go beyond conventional offerings, such as fans’ engagement.
throug live streaming capabilities [5]. A sample of these offerings is denoted from the outcomes of the 5G stadium trials that were conducted from the partnership of Nokia and Qualcomm in the World Cup at St Petersburg’s Kresovsky 5G Stadium [6]. Consequently, 5G infrastructure was sufficient enough to provide a high-fidelity video and immersive Augmented/Virtual Reality (AR/VR) experience. Ubiquitous ultra high definition and high dynamic range streams and VR can motivate an incredible shift on how sport events, for example, are enjoyed by the end user making use of 5G infrastructure [7]. A recent stadium showcase was the collaboration of Samsung and Verizon with a live 5G demonstration of streamed 4K video via a 180-degree camera, international video call and immersive VR [8].

Apart from the private sector engagement there is a significant work related to the integration of immersive media services with Software Defined Networking (SDN) and Network Function Virtualization (NFV) technologies also through European funded projects. Some examples of it are the 5GMEDIA, the 5G-Xcast project, or the Flame projects. 5G-MEDIA is a 5G PPP project, which applies SDN and NFV technologies to media applications to build an integrated programmable service platform for facilitating, designing, and developing media services over 5G infrastructures [9]. 5G-Xcast is also a 5GPPP project that demonstrates an architecture for broadcast and multicast communication over the 5G wireless networks [10]. It converts 5G infrastructure to audio-visual media content for fixed and mobile access networks, including terrestrial broadcast. FLAME project develops a platform for large-scale experimentation of future Internet media services fully integrated with wide casting supporting high mobility scenarios [11].

There are also other works focused on exploring the immersive media industry, for example, the authors in [13] the authors in [13] define an early use case where consumers can experience more than 100 Mbps data rates in the target region while enjoying facilities that require exceptionally elevated data rates. Additionally, in [14] the authors present a system architecture for live immersive 3D-media transcoding service, focusing mainly to the Quality of Experience of the users. Comparing 5GTANGO work with the above-mentioned, 5GTANGO [12] goes one step beyond by exploring the benefit of efficient monitoring mechanisms, QoS assurance and SLA management, related to the QoS demands of immersive multimedia services, running on top of the virtualized SONATA SP. Therefore, the QoS demand of the media services was investigated. The QoS level provided in 5G systems should meet the requirements of future Internet industry and go beyond what can today be accomplished with any wireless communication technology [15]. A preliminary description of the work described in this paper can be found in [16], where the authors presented an approach for mapping the high-level NS requirements to low-level network parameters. At the same time proposed a mechanism for suggesting the most important QoS parameters to the service provider. In order to achieve the best possible QoS, there is a need for advanced network monitoring for efficient resources management. Several works address this domain. Authors in [17], presented an extension of Prometheus.io, a monitoring framework implemented and integrated within SONATA project, of which 5GTANGO is an extension. In short, the SONATA monitoring framework gathers and processes data from many sources, enabling the developer to activate measurements and thresholds to capture generic or service-specific behaviors. In addition, the developer can define rules based on metrics collected from one or more VNFs in one or more Network Functions Virtualization Infrastructure (NFVIs) to receive run-time notifications.

III. IMMERSIVE MEDIA SERVICE DESCRIPTION

A. Network Service Architecture

The architecture of the service in this paper is following the cloud-native paradigm. Specifically, Continuous Integration/Continuous Delivery (CI/CD) approach has been applied during the service components development, adhering to the overall 5GTANGO Software Development and IT Operations (DevOps) approach [18]. All the components of the service are containerized and the functions are split into microservices facilitating their deployment, updating or scaling, without affecting the rest of the service components. In line with the previous statement, the VNFs are represented by a POD (a group of one or more Kubernetes containers, with shared storage/network, and a specification for how to run the containers) as illustrated in Fig.1. The subsections that follow present in more detail the NS, the interconnections between the VNFs and the service possible use cases.

B. Network Service Decomposition in VNFs

The immersive media network service is composed of several interconnected VNFs:

- **Content Manager Service (CMS):** it is the entry-point of the immersive media service. The VNF gives the possibility to register new cameras, connect new Media Streaming Engines (MSEs) and also provides the list of the available videos to the final users through a RESTful API.

- **Media Aggregator (MA):** receives the Real-Time Messaging Protocol (RTMP) video streams from the cameras, and also redirects those videos to the different MSEs or to other MAs in another network node, if necessary. The MA has a RESTful API to interact with the CMS and the monitoring system.

- **Media Streaming Engine (MSE):** this VNF receives the RTMP video flow from the different MAs and implements an adaptive streaming algorithm. Firstly, the input video is transcoded to the different qualities and then an HTTP Live Streaming (HLS) playlist is
generated with the different video segments. The MSE contains also a simple RESTful API to generate the VNF statistics and to send them to the monitoring system. This VNF is deployed on the edge of the network in order to bring the heavy transcoding task near to the user.

C. VNFs Inter-networking

It is worth mentioning that all chained VNFs of the immersive media service are deployed in separate containers, each of them realizing specific functions. Within a POD, the communication between containers is done through the loop-back interface, by specifying the port of the container (e.g. the API is assigned to port 5000). In order to access the different containers of a POD from the outside, it is necessary to use the external IP of the POD and the port of the container. The communication between the different PODs is done through the weave network plugin. This Kubernetes plugin creates a virtual network that connects the Docker containers across multiple hosts and enables their automatic discovery. In the case that there are several instances of the MA or the MSE, all instances of the same pod share the same IP address and are placed behind a load-balancer who decides the traffic steering according to set policies.

D. Data flows and Service Chaining

In the immersive media service, there are mainly two types of traffic flows, namely a) video traffic and b) control traffic. The previous mentioned flows could be either correspond to internal traffic between the different VNFs or external between the NS and the users.

- **Video Traffic:** The video traffic is the most important one in the network service, as it is the basis of the service and also because it generates a big load of information. The video traffic can be sub-divided in two types, the RTMP flow and the HTTP traffic.
  - RTMP: the RTMP video traffic is present in the links between the cameras and the MA and also between the MA and MSE. This protocol is a standard in live video streaming.
  - HTTP: in the MSE, the RTMP input is transcoded and adapted to the HLS protocol. This standard is an adaptive streaming protocol based on HTTP and the most widespread adaptive streaming protocol together with DASH. This HTTP video traffic is present between the MSE and the end-user.

- **Control traffic:** The rest of the traffic is related to the service API calls or other messages related to the service orchestration.
  - API: this RESTful API messages are used to configure the service (between the different VNFs) or contains monitoring information (inside the same pod).
  - Rest of the traffic: traffic related to message exchange with the SP for service management purposes, like instantiation and scaling.

E. Immersive Media Service Use Cases

In this sub-section the different use case scenarios of the immersive media NS are investigated. To begin with, the NS is invoked when a user logs in through one of the players (i.e. the Android player, the VR player or the web player) developed specifically for the immersive media service, in order to view the streams. The log-in of the user instantiates the various Virtual Network Functions (VNFs) that are required to run the chained NS by the 5GTANGO SP. Moving forward, three different use cases of the NS dealing with different configurations and interconnections of the internal VNFs are presented below.

- **Single camera streaming to a single/multiple users:** This use case demonstrates a single stream of video to one or multiple users the with the ability to add social networks to the stream. The NS is comprised by three VNFs, namely a) Media Aggregator (MA), b) Content Management System (CMS) and c) Media Streaming Engine (MSE).

- **Multi-location streaming to single or many users:** This use case involves streaming from multiple cameras to users. The users have the ability to control the source camera as they desire. This use case involves multiple instantiations of each of the previous mentioned VNFs (i.e. MSE, CMS, MA) based on the requirements of the NS.

- **Single location streaming with near real-time video analytics:** This use case deals with the application of real-time video analytics during the consumption of the stream. The use case can provide the user’s analytics related to the streaming video, as well as other additional info on top of the video that will enhance the immersiveness of the experience. A new Video Analysis VNF also, that is instantiated as soon as the user chooses to get the analysis of the stream they are watching is needed to be introduced in this use case.

IV. PROPOSED CAPABILITIES

A. Service Platform Overview

The main goal of the 5GTANGO is built upon the existing SONATA, a NFV framework that provides a programming model and development tool-chain for virtualized services, fully integrated with a DevOps-enabled service platform and orchestration system [25]. The SP enhances the SONATA functionality in several ways bringing forward an SLA Manager, as well as increasing the abilities of the Monitoring Manager and other components previously developed. SLAs and the suitable monitoring functionalities of the promised QoS levels, taking into consideration the available resources and the ability to ensure certain quality. Furthermore, the unified portal a easy to use interface to develop, deploy, manage and monitor a NSs.

B. SLA Management Framework

Nowadays, media applications pose specific network demands in order to deliver seamless operation to the end users. This applies especially to immersive media and live streaming
maximum values and thresholds allowed for a set of network specific Service Level Objectives (SLOs), which defines the operator is willing to take. Each SLA-T includes a set of NS, while it also describes what type of QoS commitments are needed in order to review the network’s health [21]. As stated by many service providers, they are tempted by the promise of QoS such as high speed/performance, high reliability, low latency, availability and connectivity, as well as dynamic bandwidth allocation from 5G core networks [22]. Moreover, VR/AR will be enhanced, by providing efficient SLAs in order to guarantee high definition and real-time streaming videos.

At this point a question arises: “What is actually an SLA?”. An SLA is a contract between the service provider and the end-user, which underlines the performance standards that are going to be met by the service provider [23]. In order to fulfill the gap between service providers and the end-users, 5GTANGO is a front-runner in defining SLA management mechanisms for 5G environments, by introducing a complete SLA Management Framework. The SLA Management Framework in 5GTANGO allows the end-to-end management of SLAs, starting from the template formulation to the agreement violation, considering the whole NS life-cycle. The SLA management framework is a micro-service enabled component that can be adapted and extended to work on different 5G-oriented service platforms, supporting a variety of RESTful APIs. The workflow of the SLA Management Framework is partitioned into two phases, namely a) SLA Template Management and b) the Information Management [16]. The first phase takes place prior to the NS deployment, includes the template formulation, while the second phase starts during the NS deployment, incorporating the SLA instance creation with the violations management, as depicted in Figure 2.

1) SLA Template Creation: Firstly, the service provider defines an initial SLA Template (SLA-T). The SLA-T refers to an initial advertisement of the provider regarding the attached NS, while it also describes what type of QoS commitments the operator is willing to take. Each SLA-T includes a set of specific Service Level Objectives (SLOs), which defines the maximum values and thresholds allowed for a set of network services, like the presented one, that benefit from technologies like AR/VR [20]. 5G environments are expected to unlock the full potential of these media services, where QoS assurances through SLA enforcement is a critical process, and monitoring information is needed in order to review the network’s health. As stated by many service providers, they are tempted by the promise of QoS such as high speed/performance, high reliability, low latency, availability and connectivity, as well as dynamic bandwidth allocation from 5G core networks [22]. Moreover, VR/AR will be enhanced, by providing efficient SLAs in order to guarantee high definition and real-time streaming videos.

At this point a question arises: “What is actually an SLA?”. An SLA is a contract between the service provider and the end-user, which underlines the performance standards that are going to be met by the service provider [23]. In order to fulfill the gap between service providers and the end-users, 5GTANGO is a front-runner in defining SLA management mechanisms for 5G environments, by introducing a complete SLA Management Framework. The SLA Management Framework in 5GTANGO allows the end-to-end management of SLAs, starting from the template formulation to the agreement violation, considering the whole NS life-cycle. The SLA management framework is a micro-service enabled component that can be adapted and extended to work on different 5G-oriented service platforms, supporting a variety of RESTful APIs. The workflow of the SLA Management Framework is partitioned into two phases, namely a) SLA Template Management and b) the Information Management [16]. The first phase takes place prior to the NS deployment, includes the template formulation, while the second phase starts during the NS deployment, incorporating the SLA instance creation with the violations management, as depicted in Figure 2.

1) SLA Template Creation: Firstly, the service provider defines an initial SLA Template (SLA-T). The SLA-T refers to an initial advertisement of the provider regarding the attached NS, while it also describes what type of QoS commitments the operator is willing to take. Each SLA-T includes a set of specific Service Level Objectives (SLOs), which defines the maximum values and thresholds allowed for a set of network services, like the presented one, that benefit from technologies like AR/VR [20]. 5G environments are expected to unlock the full potential of these media services, where QoS assurances through SLA enforcement is a critical process, and monitoring information is needed in order to review the network’s health [21]. As stated by many service providers, they are tempted by the promise of QoS such as high speed/performance, high reliability, low latency, availability and connectivity, as well as dynamic bandwidth allocation from 5G core networks [22]. Moreover, VR/AR will be enhanced, by providing efficient SLAs in order to guarantee high definition and real-time streaming videos.

At this point a question arises: “What is actually an SLA?”. An SLA is a contract between the service provider and the end-user, which underlines the performance standards that are going to be met by the service provider [23]. In order to fulfill the gap between service providers and the end-users, 5GTANGO is a front-runner in defining SLA management mechanisms for 5G environments, by introducing a complete SLA Management Framework. The SLA Management Framework in 5GTANGO allows the end-to-end management of SLAs, starting from the template formulation to the agreement violation, considering the whole NS life-cycle. The SLA management framework is a micro-service enabled component that can be adapted and extended to work on different 5G-oriented service platforms, supporting a variety of RESTful APIs. The workflow of the SLA Management Framework is partitioned into two phases, namely a) SLA Template Management and b) the Information Management [16]. The first phase takes place prior to the NS deployment, includes the template formulation, while the second phase starts during the NS deployment, incorporating the SLA instance creation with the violations management, as depicted in Figure 2.

1) SLA Template Creation: Firstly, the service provider defines an initial SLA Template (SLA-T). The SLA-T refers to an initial advertisement of the provider regarding the attached NS, while it also describes what type of QoS commitments the operator is willing to take. Each SLA-T includes a set of specific Service Level Objectives (SLOs), which defines the maximum values and thresholds allowed for a set of network

2) SLA Instance Creation: The SLA Instance (SLA-I) is the enforced SLA-T (i.e. final agreement between service provider and customer) with the instantiation information of an SLA-T, that refers to the linked NS instance and the customer’s information. The SLA-I is automatically created after the successful NS instance creation, and thus the Monitoring Manager should be informed in order to provide monitoring data to the SLA Manager for the guaranteed SLOs. To this end, when the NS is successfully instantiated, the appropriate monitoring rules are formulated and sent towards the Monitoring Manager. More details for the monitoring management are given in the sub-section IV-D.

3) SLA Violation: During this stage, the SLA Manager is responsible to identify any violations occurred and take the necessary actions. Once a metric exceeds the guaranteed threshold, the Monitoring Manager, produces an alert (i.e. sub-section IV-D-5). As soon as the the SLA Manager consumes the alert, the overall value of that specific metric is calculated and is decided whether the corresponding SLA-I is violated, or not. The violation data is stored in an internal database in order to be available for visualization in a unified portal. A detailed description of an end-to-end work-flow is described in [24].

C. Live Streaming QoS Demands

The provided SLAs aims to guarantee the desired QoS of the provided immersive media service. QoS requires the examination of the overall traffic requirements, as well as examined in the different interconnected VNFs. To achieve the quality of a live streaming video it should be taken in mind not only the network elements but also configurations and coordination at the endpoints. In general, the quality of the video implies:

- Optimization of the endpoint configurations (e.g., resolution, frames per second)
- Optimization of the transport layer - Optimize network to transfer video traffic across network
- Interoperability components - Video clips often include endpoints of varied capabilities. The design and configuration of systems to maximize interoperability can affect the quality of the video

In the current solution of the described media service, different QoS parameters are considered and tested in order to evaluate the implemented streaming NS, in three discrete levels: “Gold”, “Silver” and “Bronze”. For each SLA level a set of specific SLOs were considered as depicted in I. The above-mentioned media quality levels are effectively monitored, by providing monitoring rules towards the Monitoring Framework, in order to be measured and analyzed, for any violation detection. This interaction between the SLA Management and Monitoring Framework can eventually save precious time and money to service providers.
TABLE I: Defined SLA levels for the immersive media service.

<table>
<thead>
<tr>
<th>SLA</th>
<th>Availability</th>
<th>Bandwidth</th>
<th>Jitter</th>
<th># input-output connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOLD</td>
<td>99.95 %</td>
<td>&gt;= 80Mbps</td>
<td>&lt; 20ms</td>
<td>&lt;= 100</td>
</tr>
<tr>
<td>SILVER</td>
<td>98 %</td>
<td>&gt;= 65Mbps</td>
<td>&lt; 35ms</td>
<td>&lt;= 80</td>
</tr>
<tr>
<td>BRONZE</td>
<td>95 %</td>
<td>&gt;= 50Mbps</td>
<td>&lt; 50ms</td>
<td>&lt;= 65</td>
</tr>
</tbody>
</table>

D. Monitoring Management Framework

Current available monitoring tools, frameworks and platforms fall short in satisfying an efficient way the requirements of a 5G network. The involved stakeholders in many cases are: a) intrusive and heavy-handed for short-lived, lightweight cloud instances; b) not covering deployments in both hyper-visor-based and containerized manner; c) not able to follow the fast pace of deployment changes enforced by continuous dynamic scheduling, provisioning and auto-scaling; and d) they are bounded to specific cloud platform technologies, making them inappropriate for open environments, such as the one envisioned in 5GTANGO. Under this prism, the development of a monitoring framework is a necessary but not trivial task in today’s complex and heterogeneous infrastructure environments, given the applications diversity that are instantiated therein. With these challenges in mind, the Monitoring Framework, as developed and integrated within the Service Platform of 5GTANGO project includes several components and provides a wealth of functionalities to fulfill the requirements of all involved stakeholders. A high-level architecture diagram is shown in Figure 3.

In brief, the 5GTANGO Monitoring Framework [17] collects and processes data from several sources, providing the developer the ability to activate metrics and thresholds in order to capture generic or service-specific behaviour and fulfill the guaranteed SLAs. Moreover, the developer can define rules based on metrics gathered from one or more VNFs (either in virtual machines or containers) deployed in one or more NFVIs in order to receive notifications in real time. The next sub-sections provide an overview of the main functionalities that the Monitoring Framework provides.

1) Collecting Data From Several Sources: It is of paramount importance to collect monitoring data from as many as possible sources. In the implemented framework, the following types of sources for collecting data are available:

- **Network infrastructure monitoring:** data that is accessible from physical or virtual network elements or through the Network Management System (NMS);
- **NFV infrastructure monitoring:** usually implemented with OpenStack, the data collection methods go beyond the native Ceilometer service - for better granularity, the open-source software Netdata will be exploited;
- **Kubernetes infrastructure monitoring:** monitoring data will be collected per cluster node, per POD and per container using cloud native tools already available;
- **VNF monitoring:** data collected with monitoring probes from the VNF specific monitoring metrics, exposed by each VNF according to the developer-provided descriptor;
- **SDN monitoring:** data collected from the OpenDayLight controller, with respect to the control and data plane of the network;
- **MANO monitoring:** data collected from network services and functions, integrated with V&V through the appropriate APIs.

2) Prometheus: 5GTANGO adopted Prometheus monitoring tool well before it became a Cloud Native Computing Foundation (CNCF) project and properly enhanced its functionalities to address additional 5GTANGO requirements. Prometheus is an open-source service monitoring system, based on time series database, that implements a highly dimensional data model. Prometheus is responsible for collecting the data and communicating with the time-series database for retrieving data upon request.

3) Push Gateway: This component is part of the open source Prometheus monitoring solution that has been adopted and extended to cover the needs of 5GTANGO Monitoring Framework. Push Gateway is utilized so that the exporters/sources use HTTP POST method to “push” monitoring data to the Push Gateway and towards other components like the SLA Manager, while Prometheus collects the data in a predefined time interval. The advantage of this approach is that in the case of the deployment of a new service, there is no need for Prometheus to search for data related to the newly deployed VNF, but rather collect them from the Push Gateway.

4) SNMP Manager: 5GTANGO Monitoring Framework provides the ability to collect data from VNFs using Simple Network Management Protocol (SNMP) protocol. This functionality increases the flexibility of the monitoring framework taking into account that many VNF developers use SNMP, as a standard and widely accepted approach, to provide metrics that are related with the performance of their services. Currently, implementation supports public access via SNMPv2 and also secured access via SNMPv3. In order to enable the SNMP functionality, the developer must provide the necessary information (metrics, SNMP OIDs, time interval, etc.) inside the VNF descriptor. For this purpose, the VNF descriptor schema of 5GTANGO has been extended accordingly.

5) Alert Manager: The Alert Manager is responsible for sending notifications about firing alerts to the subscribed users. After this notification, the user can take advantage of the Monitoring Manager API to further investigate the fault by retrieving past monitoring data. Regarding the QoS alerts, SLA
monitoring rules and SLA violations are produced also in form of notification alerts. Thus, the message (i.e. alert) is pushed to the SLA Manager for managing the SLA violations.

6) Monitoring Manager: Monitoring manager offers RESTful APIs to the users with respect to the monitoring data of their instantiated 5G services, including: 1) the relation among services, network functions, NFVIs and users, 2) the ability to modify rules and thresholds during service/function runtime, 3) the reconfiguration of Prometheus, and 4) the capability to provide information on collected monitoring metrics.

V. Evaluation Results

In order to evaluate the performance of the proposed immersive media NS in terms of efficiency and ease of use, the proposed architecture was included in the innovative 5G infrastructure environment of the 5GTANGO Service Platform. Specially, the SLA Management and Monitoring Management Frameworks were adopted to enable the enhancement of the scenario for live streaming content using VNF features of the SP. In order to check the performance of the service and to validate the SLA levels defined in the Table I, two different test-cases have been carried out on a single network service deployed on the SONATA powered by 5GTANGO SP. The first test-case using different input combinations on the MA and the second one with different number of clients consuming the videos from the MSE. Each test-case was repeated twenty times (each repetition running for 72 hours) and the outcomes were extracted and averaged to produce statistical evaluation results. During these tests, interesting metrics of the MSE have been extracted (i.e. due to the fact that this is the component that requires more hardware resources for the video transcoding process) as well as some important metrics about the service (i.e. availability, video bit-rate and number of connections).

The first test-case uses three different combinations of inputs to the MA: a) One 360 video (equirectangular projection, 2880x1440 pixels) plus a plane video (1920x1080 pixels); b) Two 360 videos plus two plane videos and c) two 360 videos plus three plane videos. The hardware resources that were used and the bit-rate generated in the MSE output are depicted in Figure 4 for each input combination. Note the increase in CPU utilization by increasing the number of 360 videos as well as the bit-rate.

The second test-case is related to the generation of two inputs on the MA, using a 360 video plus a plane video, and simulate different number of clients connected to the MSE. The hardware resources used and the bit-rate generated are shown in Figure 5. This test-case shows that the number of clients connected to the service hardly causes variations in the consumption of MSE hardware resources, thus achieving a system suitable for a greater number of users.

In terms of availability, the service which was used to make these tests has been up and running for 72 hours and the total measured downtime of the service was around 5 minutes. This translates into 99.99 % of service availability, with only 0.001 % of downtime. Comparing the results with the defined SLOs, the service complies with all the SLAs levels. After taking some bit-rate measurements with all the actors (i.e. service components and users) in the same network, we obtained an average result of 89Mbps. With these results, the SLOs about bandwidth are complied under this use case. The jitter values proposed on the SLOs are met in a considerable manner between the MA and a user for the three SLAs levels. Regarding the number of input and output connections, the service supports in the MA input at the same time up to two 360 degree video streams and up to three plane Full-HD videos. Also a single instance of the MSE supports up to 120 clients consuming the HLS playlist and video fragments. This number of connections complies with all the defined SLA levels.

VI. Conclusions

In this paper, we have introduced a novel architecture of a live streaming NS deployed in a real 5G Platform - the NFV-enabled SONATA powered by 5GTANGO SP. The results of our experiments indicate several insights in regards of the service’s behaviour. First of all, we have successfully fulfilled the signed SLOs in the corresponding SLA, proving that the promised quality of service by the service provider is feasible. Later, the results of the experiments justified the fact that it is better to deploy the MSE VNF in the Edge due to its hardware requirements and also not to saturate the core of the network with complex video processing. It is important to highlight that the video transcoding does not depend on the number of
clients because the process is done only one time, but the bit-rate generated is completely related with the number of videos and the number of clients.

Future work may include the extension of the evaluation of the provided media service by testing it with additional hardware (i.e. more cameras, different types, etc.), data-sets of different sizes and more distributed set of end-users. Apart from this, we also aim to extend the whole architecture of the service in order to enable Quality of Experience (QoE) enforcement, considering parameters based on real-time monitoring feedback. Additionally, the current architecture of SONATA is able to monitor and manage business guarantees in a 5G environment. Therefore, we envision to manage a 5G network with multiple domains, enabling higher level of integration and at the same time, to adapt the proposed service architecture from a wide range of media verticals, allowing higher level of abstraction.

ACKNOWLEDGMENT

This work has been partially supported by the 5GTANGO project, funded by the European Commission under Grant number H2020ICT-2016-2 761493 through the Horizon 2020 and 5G-PPP programs (http://5gtango.eu).

REFERENCES