A System Architecture for Live Immersive 3D-Media Transcoding over 5G Networks

Alexandros Doumanoglou
ITI/CERTH, Greece
aldoum@iti.gr

Nikolaos Zioulis
ITI/CERTH, Greece
nzioulis@iti.gr

David Griffin
UCL, UK
d.griffin@ucl.ac.uk

Javier Serrano
UPM, Spain
jsr@gatv.ssr.upm.es

Truong Khoa Phan
UCL, UK
t.phan@ucl.ac.uk

David Jiménez
UPM, Spain
djb@gatv.ssr.upm.es

Dimitrios Zarpalas
ITI/CERTH, Greece
zarpalas@iti.gr

Federico Álvarez
UPM, Spain
federico.alvarez@upm.es

Miguel Rio
UCL, UK
miguel.rio@ucl.ac.uk

Petros Daras
ITI/CERTH, Greece
daras@iti.gr

***Abstract*—The upcoming 5G networks, among other technological advances, bring Network Function Virtualization (NFV) capabilities enabling deployment of application service intelligence on their Next Generation Core (NGC). Application specific logic is packaged into Virtual Network Functions (VNFs) so that their instantiation and deployment can be done at any node of the NGC, with their management and orchestration being maintained by the 5G infrastructure. While the number of instances of each VNF and their placement inside the NGC network are managed by the 5G infrastructure, such management cannot be optimal without application context. In this paper, we propose a 5G oriented system architecture for a next generation augmented virtuality tele-immersive two-player video game application. In the presented video game, the players compete in a capture the flag race in an innovative game movement control setting which uses motion capture technology to allow the players to interact with the game via their body posture and hand gestures. On the top of this, real-time 3D-Reconstruction technology is utilized to create 3D avatars of the players and embed them inside the game environment. Apart from the players, the application also supports real-time spectating of the game action by a considerable amount of spectators that join the live game via client software designed for desktop PCs, smartphones and tablets, connected through mobile or fixed access networks. To distribute the 3D traffic to such a number of consumers that have different device capabilities and are located at varying geographical locations while offering the highest possible Quality of Experience (QoE), is a challenging task. One of the contemporary ways to address this problem is via adaptive streaming. To realize this concept, real-time 3D-Media Transcoders need to be employed. The proposed system architecture considers packaging the aforementioned 3D-Media Transcoders as VNFs that can be deployed on 5G infrastructure. In the paper, it is shown that such an architecture can decrease costs for a given level of offered QoE, with evident benefits for the game service's shareholders. While the application type presented in this paper is fixed, the proposed system architecture can be adopted by other applications of similar context with similar benefits gained from the flexible deployment of virtualised applications in 5G networks.**

I. Introduction

Augmented Virtuality (AV) is a subcategory of Mixed Reality (MR) which refers to the merging of real world objects into virtual worlds. This merging is oftentimes accomplished via 3D-digitizing the real world objects and rendering them inside the Virtual Environment (VE) [1]. Modern Tele-Immersion (TI) platforms [2] are actually immersive AV applications that 3D reconstruct humans in real-time and embed their avatars inside a shared VE, where they can real-time interact with the elements of the VE and the other participants. Typically, AV TI applications produce large volumes of visual data in the form of watertight textured 3D meshes, thus, creating a challenging networking scenario. In order to cope with this amount of data produced by a TI application, the invention of novel, efficient, real-time, and potentially adaptive, 3D compression schemes is highly demanded. However, due to the difficulties of developing such decent textured 3D mesh codecs conforming to the aforementioned characteristics, an upgrade to a more efficient network infrastructure appears to be an additional necessity. TI applications have very high bandwidth and ultra-low latency requirements, in order to allow real-time interactivity. Coincidentally, these requirements comprise some of the main targets to be addressed by modern 5G networks.

In order to address all those targets, among other technological advances, the new 5G networks will introduce Mobile Edge Computing (MEC) [3] capabilities enabled by Network Functions Virtualization (NFV) and Software Defined Networking (SDN). With the help of NFV and SDN, the embedded resources at the mobile network edge are employed to offer added value services, improve Quality of Experience (QoE) by moving intelligence at the edge and create new business opportunities. In particular, the virtualization capabilities offered by NFV and SDN will help media service providers to exploit resources from a central point, without being worried about the location of the actual hardware, its maintenance and its vendor. Moreover, orchestration capacity makes it possible to coordinate thousands of resources, like unconventional hardware, e.g. GPUs, in fly and on demand. Programmability and automation are two important features which change the overall media service behavior towards more automated/intelligent systems able to guarantee QoS and SLA compliance with minimum human intervention and error. Dynamic scaling is another important advantage enabled by SDN/NFV to optimize resource utilization and reduce OPEX. Furthermore, sharing expensive infrastructure among

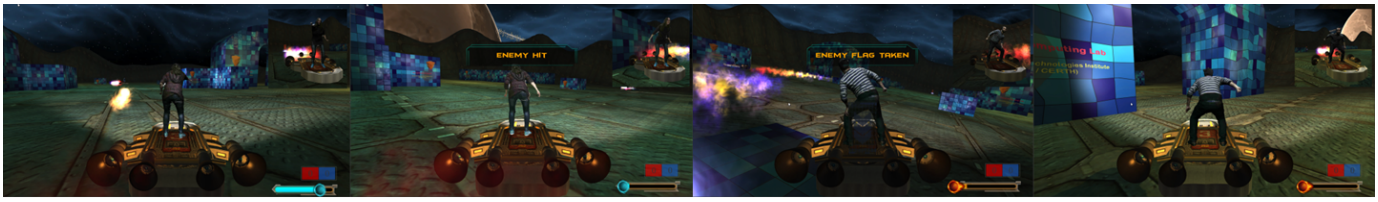

Fig. 1. Screenshots of the video game taken from the players' clients.

serves tenants/service provides, so called 5G multi tenancy feature, will significantly reduce the required CAPEX for service offering. Last but not least, SDN/NFV create an open ecosystem where a full choice of modular plug-ins can be easily adapted to customize service offering according to the user needs.

In this paper, we consider a realization of a TI application over the new 5G-network virtualization technologies. The TI application presented in this work, constitutes a competitive two-player AV Tele-Immersive video game in which, the players' 3D avatars are embedded inside the game environment. Furthermore, the application allows for live spectating of the gaming sessions by a paramount number of interested users distributed around the globe. Given the fact the the textured 3D meshes corresponding to the players' avatars provoke an increased amount of network traffic to be circulated among the participants, we propose that 3D media adaptive streaming can be utilized to stabilize the QoE of the participating parties and confront QoE drops due to network congestion. Within the context of 5G, this scenario can be realized via a network aware media application that implements a real-time adaptive streaming media delivery framework via Virtual Network Functions (VNFs). The application-specific VNF that is presented in this paper (vTranscoder) is instantiated at the 5G Next Generation Core (NGC) [4] and is responsible for transcoding the textured 3D mesh of each player to varying levels of quality in real-time to support adaptive streaming for players and spectators.

The contributions of this paper is twofold. First, it is one of the pioneering works that studies AV Tele-Immersive applications in a 5G network setting and second, it proposes a transcoder VNF to formulate a novel 5G-oriented media application architecture, potentially improving the QoE of the application's users and decreasing the costs for the application's shareholders.

II. Scenario

The envisaged scenario of this paper is the realization of a next-generation AV Tele-Immersive video game. In this video game, two players combat for their flags in a typical capture-the-flag setting. Each one of them has the goal to capture the opponent's flag from their base and bring it safely on their own base to score a point. Both players are located in special capturing stations equipped with proper hardware and software and they are 3D reconstructed in real-time. Their 3D-reconstructed avatars are embedded inside the game environment and placed over sci-fi hover-boards. The players are able to navigate inside the game world by using their body posture. Bending their knees makes them move forward, while leaning their torso left or right allows them to make turns. Capturing the opponent's flag is accomplished by the player moving over it. Furthermore, the players are given the opportunity to release fireballs to each other by using a special gesture of their right arm. By positioning their right arm straight backwards, they make the fireball "charge" while stretching their arm forward is releasing the fireball towards their opponent. When a player is hit by a fireball a visual effect is triggered that breaks their avatar into pieces, serving as a visual feedback for this action. The consequence of the hit is that the hit player is forced to start from scratch: i.e. the captured flag is returned to their opponent's base and their position is reset. A more detailed description of this video game can be found in [5], while screenshots are illustrated in Figure 1.

Apart from the players participating in the game, the software allows other participants to join as "spectators". The spectators can join a live game and watch the action in real-time. Special spectator software is available for different types of devices: desktop PCs, phones and tablets. For desktop PCs a standard keyboard mouse interface is given, while for phones and tablets an Augmented Reality (AR) mode is available that permits the spectators to view the game action from arbitrary viewpoints on a registered planar surface of the real world.

The realization of the aforementioned scenario has to face the following challenges:

- Supporting a considerable number of spectators, joining from different parts of the globe.
- Estimate QoE of the players and the spectators by collecting live metrics of the gaming session.
- Maximize the spectators' and players' QoE, by supporting real-time live adaptive streaming of the players' 3D avatar streams. This is different than typical pre-stored and pre-encoded content found in standard cases of on demand video streaming. Contrariwise, in order to support live adaptive streaming of the players' 3D avatar streams, a real-time textured 3D mesh transcoding service is required.
- Minimize the latency between the players' 3D appearance streams and the game's state by controlling computational and network resource utilization. Such a latency minimization contributes to higher QoE for players and spectators.

- Optimize network and computational resource utilization and placement while maintaining a high QoE for spectators and players. Such an optimization reduces costs for the application's shareholders.

III. System Architecture

In this section, we present a network-aware system architecture that covers the realization of the scenario described in Section II. The aim of the section is to showcase how modern demanding network applications can benefit from the technological advances being brought by the upcoming 5G networks. We begin by introducing the application's components. Subsequently, we give an outline about the placement of the components and the data flow. We continue by giving detailed information about the 3D-Media transcoding service that enables live 3D-Media adaptive streaming and finally, elaborate on the deployment of the application on 5G NGC.

A. Application Components

The AV Tele-Immersive video game application considered in this paper consists of the following software components:

1) A dedicated 3D reconstruction component (referred as *"3D-Reconstructor"*) instantiated once for each player at the corresponding capturing station with the responsibility to gather RGB-D frames from depth sensors and fuse their data to produce a textured 3D-Mesh representation of the player. This representation is later embedded inside the game environment.
2) A body-pose recognition component (referred as *"Body-pose recognizer"*) with the responsibility to provide 3D skeletal information about the captured player using motion capture technology.
3) A dedicated game server component that is responsible to synchronize the game state across players' and spectators' clients.
4) A player client component, instantiated once for each player, with twofold responsibility. First, to translate the skeletal information provided by the body pose recognition component to game specific commands, allowing the players to navigate inside the virtual environment and release fireballs to their opponents. Second, to visualize the current game state with the additional embedding of the 3D reconstructed avatars of the players inside the game world.
5) A desktop PC spectator client component which can join a live game being played in a game server and enables live free viewpoint spectating of the game action.
6) A spectator client component designed for smart-phones with functionality similar to the corresponding desktop component. This software component is properly optimized for mobile screens.
7) A spectator client component designed for tablets that additionally allows spectating live game action in AR mode.
8) Dedicated 3D Transcoder components (referred as *3D-Media transcoders*), that transcode the textured 3D reconstructed Meshes of the players in various qualities in real-time to allow live adaptive streaming of the 3D content to the players' and the spectators' clients.

B. Software component placement and data flow

Depending on the different aspects of the application's scalability, various component placement strategies may be employed. In the present paper, we mainly focus on a strategy that will enable a considerable amount of spectators to join a single game session and from any device. The placement of the software components are presented first and the data flow between the components is given second.

1) Component Placement:

- **3D-Reconstructors**: They are deployed on dedicated PCs located at the players' capture stations.
- **Body-Pose Recognizers**: Similar to 3D-Reconstructors, they are deployed on PCs located at the players' capture stations.
- **Game Server**: The game server is considered to be deployed on the application shareholders' premises.
- **Player Clients**: They are deployed on PCs located at the players' capture stations.
- **Spectator Clients**: They are deployed on PCs/phones/tablets located anywhere.
- **3D-Media Transcoders**: They are deployed as VNFs on the 5G network NGC infrastructure and their instantiation and placement is managed by the 5G network's management and orchestration module, based on hints provided by the application.

2) Data flow: The *logical* data flow communications between the different application's components are as follows:

- **Local player's *3D-Reconstructor*** $\rightarrow$ **Local *Player Client***: The 3D avatar representation of the player that was reconstructed at the capture station is transmitted to the local player's client for the visualization of self-representation inside the game.
- ***3D-Reconstructor*** $\rightarrow$ ***3D-Media Transcoders***: The 3D avatar representation of the players, that were reconstructed in the capture sites, are transmitted to the 3D-Media Transcoders that take over the transcoding action enabling adaptive streaming for the remote player and the spectators.
- **Local player's *Body-pose recognizer*** $\rightarrow$ **Local *Player Client***: The 3D skeletal information produced by the *Body-pose recognizer* is transmitted to the local player's client in order to translate it to game specific commands.
- ***Player Clients*** $\leftrightarrow$ ***Game Server***: A bidirectional data flow is established between the player clients and the game server for synchronizing the players' game commands and overall game state.
- ***Game Server*** $\rightarrow$ ***Spectator Clients***: Spectator clients receive game state updates from the game server and visualize the game world in a consistent manner.
- ***3D-Media Transcoders*** $\rightarrow$ ***Player Clients***: Player Clients receive the 3D avatar representation of the remote player from the 3D-Media Transcoders in an adaptive way.

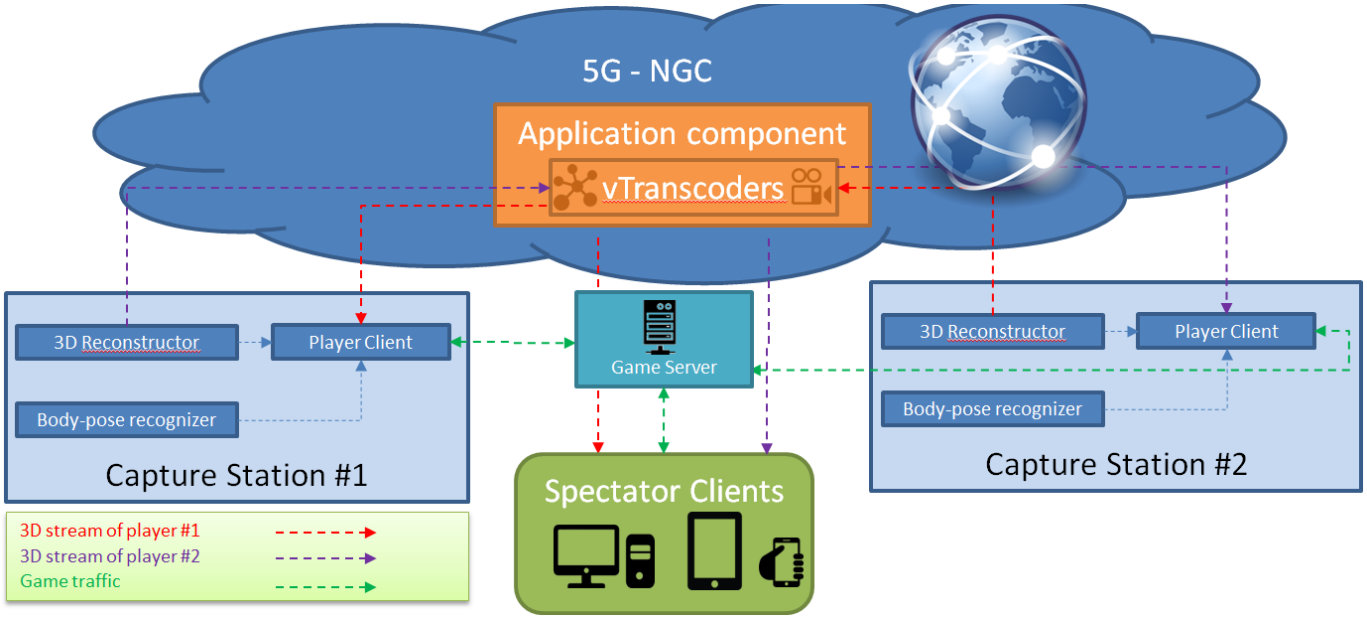

Fig. 2. Logical data-flow between software components of the studied application.

- ***3D-Media Transcoders*** → ***Spectator Clients***: Spectator Clients receive the 3D avatar representation of both players from the 3D-Media Transcoders based on the adaptation logic of the application.

In addition to the previous description, it is important to clarify that the traffic corresponding to the 3D reconstructed avatar representation of the players does not pass through the game server. This avoids the unnecessary delay that would be introduced to the 3D appearance streams by the game server when redirecting the traffic, at the cost of non explicit game-state to visual representation synchronization. Instead, the players' and spectators' clients directly receive the 3D traffic through the 5G network and not from the game server. This solution scales better as the number of spectators increases. In Figure 2 the *logical* data-flow communication between the application's software components are summarized.

C. 3D-Media Transcoders

The task of 3D-Media Transcoders is to re-encode the high quality input textured 3D mesh stream to various qualities, in real-time, in order to enable streaming adaptation for the users of the game service improving their overall QoE. Each consumer of the 3D stream (player and spectator clients) has different optimal QoE conditions, depending on device capabilities (such as device processing power in order to real-time decode the stream, screen size, etc.) and current network conditions. The input and output 3D streams of the transcoders consist of two components: 3D geometry and textures. The transcoding of the 3D stream is performed in its two components independently and various final qualities can be produced by combinations of the qualities of the individual components. The transcoding process introduces an extra delay on the propagation of the stream due to processing. However, this extra delay is compensated by the reduced size of the output data, which can propagate faster through the network. Furthermore, depending on the available hardware of the compute nodes that host the 3D-Media Transcoders, it is also possible that apart from traditional CPU-based transcoders, other implementations can utilize GPUs, to either further reduce the transcoding processing time of a single quality, or produce multiple qualities in the same unit of time.

D. 5G network deployment

The scenario and technologies proposed in this paper are focused to the efforts towards providing a SDN/NFV 5G mobile packet core [6], where components are going be deployed, with users connected through the 5G radio access network [7] or fixed access networks. When deploying the presented game service to a 5G network infrastructure, additional communication is required between the 5G network's management and orchestration service and a special software component of the application which is responsible for collecting relevant measurements and QoE metrics from the players' and spectators' clients, as well as appropriate information from the 3D-Media Transcoders.

Each of the vTranscoders will be deployed as a software VNF in one of many cloud node locations in the Network Function Virtualisation Infrastructures (NFVIs) managed by the Service Virtualisation Platform Operator which has been contracted by the TI Game Service Provider to manage the deployment of the service. Each potential cloud node for running a vTranscoder will have different characteristics in terms of its location, capabilities and costs. The potential locations include edge nodes co-located with access network functions, regional clouds operated by the 5G network operator or in central cloud locations operated by third parties. Node

characteristics vary in terms of the computational and storage capacity and whether specialised hardware such as GPUs are available. The cost of deploying a service in a particular cloud locations depend on the price offered by the cloud provider. Resource-constrained edge nodes are likely to be of higher cost than large-scale central cloud node locations.

The Platform Operator will optimise the deployment of the vTranscoders to maximise Quality of Experience within a cost budget determined by the Service Provider. A naive approach would be to deploy vTranscoders as close as possible to each of the players. This would ensure that the 3D Mesh streams are compressed as early as possible along the paths to the other players. However, the closest cloud nodes may be costly or may not have the computational resources available to compress the 3D Meshes to levels suitable for transmission over the bottleneck links. Furthermore, if there are several players in a geographical region it may be possible for them to share a vTranscoder node in a location with higher computational capacity at a lower cost than deploying individual vTranscoders for each player. Placement optimisation decisions such as these will be made by the Platform Operator according the the predicted and actual demand from players and as background network traffic levels vary reducing the throughput on some paths.

On top of this, the 5G infrastructure can additionally optimize the routing of the traffic to avoid unnecessary duplicate traffic flow across different geographical regions. This will help to establish network slicing [8] for media services when required. Deployment costs for the application's shareholders, for a target level of consumers' QoE, can also be minimized by carefully favoring spending computational resources over network resources, or vice versa, depending on all the previously described factors that are imposed by the consumers. The result is an architecture capable of providing an optimized service, considering the requirements posed and based on user perceived experience.

IV. Conclusion

In this work, a system architecture to support live 3D-Media Transcoding in the deployment of a next-generation AV Tele-Immersive video game service over 5G networks, was presented. Initially, the challenges of realizing such an application were introduced. Subsequently, all the software components of the application along with their placement and logical dataflow communications were described in detail. By leveraging the virtualization capabilities of the next generation 5G networks, we propose that the 3D-Media Transcoders are packaged into application specific VNFs with clear benefits for the application's shareholders and the end users. As instantiation and placement of the VNFs are managed by the 5G infrastructure, their scalability and placement can be optimized to offer higher QoE to the end users while reducing costs for the application's shareholders. While in this work we focused on optimizing the distribution and the quality of the 3D traffic, in the future, the 3D-Reconstructor components and the game servers could also be moved to specialized VNFs being deployed to any 5G infrastructure, further improving the quality of the game service.

Acknowledgment

This work was supported by the EC Project 5G-MEDIA (www.5gmedia.eu). This project has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement No 761699.

References

[1] [Online]. Available: https://en.wikipedia.org/wiki/Mixed_reality#Augmented_virtuality

[2] A. Karakottas, A. Papachristou, A. Doumanoglou, N. Zioulis, D. Zarpalas, and P. Daras. Augmented VR, IEEE Virtual Reality, Mar 18 - 22, 2018. [Online]. Available: https://www.youtube.com/watch?v=7O_TrhtmP5Q

[3] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54–61, 2017.

[4] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[5] N. Zioulis, D. S. Alexiadis, A. Doumanoglou, G. Louizis, K. C. Apostolakis, D. Zarpalas, and P. Daras, "3D tele-immersion platform for interactive immersive experiences between remote users," in *2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016*, 2016, pp. 365–369. [Online]. Available: https://doi.org/10.1109/ICIP.2016.7532380

[6] V.-G. Nguyen, A. Brunstrom, K.-J. Grinnemo, and J. Taheri, "Sdn/nfv-based mobile packet core network architectures: a survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1567–1602, 2017.

[7] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, "5g new radio: waveform, frame structure, multiple access, and initial access," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 64–71, 2017.

[8] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network slicing based 5g and future mobile networks: mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.