5GinFIRE test services offer to the 5G Media Vertical

Abstract

This document aims to present current and future 5G test services to be offered to the 5G Media vertical.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 732497. This publication reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Executive summary

The 5G Media Vertical Testbed offers a comprehensive set of functionalities to experimenters to enable deployment, management and monitoring of 5G-enhanced media services. Specifically, the testbed comes with two classes of applications ready for use (VR streaming and live uplink streaming), offers network assistance and edge processing to media applications, and exposes 3rd party APIs to allow application providers to monitor and manage the network slice where their services are deployed.

The work leading to the testbed, as well as its usage feedback, gave us valuable information that enabled us to identify a number of tests that need to be performed in order to verify that 5G (media slices) fulfill the needs and requirements of current and future media services.

In order to foster the adoption of 5G from the vertical sectors, collaboration with other test environments (such as 5Groningen in the Netherlands) are envisioned. Furthermore, to enable even more advanced (media) tests in the future, we plan to upgrade the 5G media vertical infrastructure with new cloud and mobile hardware (new base stations, storage units, RAMs and GPUs) and software (new 5G core) and end user devices (AR/VR goggles, haptic gloves).

List of authors

Version	Company	Author	Contribution
1.0	TNO	Lucia D'Acunto	Executive summary, chapters 1, 2, 3, 7
1.0	TNO	Piotr Zuraniewski	Chapters 4, 5, 6, 7, 8

Table of Contents

1 Introduction

Audio-visual media generates the majority of traffic today, both in mobile and fixed networks, and this traffic is expected to grow in the coming years. Additionally, new types of media services are rising, such as Virtual Reality and soon Mixed and eXtended Reality[1], which are expected to pose new stringent requirements to (mobile) networks, in order to provide a feeling of immersion and smooth interaction. At the same time, some sectors of the media industry wish to turn to mobile for the delivery of their services, most notably the (semi) professional media production sector[2].

These developments require 5G to be able to offer, on demand, capabilities for low latency and high, consistent bandwidth to the actors of this important vertical industry. The 5G Media Vertical testbed offered by 5GINFIRE addresses this challenge by showing how an orchestrated 5G network can be used to provide the required capabilities, at the hand of two relevant (future) use cases: (i) high quality live uplink video streaming, and (ii) low latency VR streaming.

2 Current technologies – State of the art

2.1 Current technologies

Online downlink media streaming services are built according to an Over-The-Top (OTT) paradigm, where the network is assumed to be unreliable and without throughput guarantees. Within this world, a lot of 'intelligence' is put within the media client that downloads the content. The media content is typically provided in different bitrates and each version is segmented in temporal chunks. The media client continuously measures the received throughput and, based on that, selects the bitrate for the next content chunk to download. This way, the client is resilient to changes in network throughput and it can guarantee continuous playback. This method of downlink media streaming is called "Adapted Bit Rate" (ABR), of which the HLS protocol from Apple[3] or the DASH standard from MPEG[4] are examples currently used in the industry. A disadvantage of this approach is that it needs to be conservative: since the network is assumed to be unreliable and best effort, this approach always reserves a 'buffer' to face abrupt network changes, which means that the streaming client often receives a stream at a bitrate (and thus at a audio-visual quality) lower than it may be necessary. Also, this approach results in relatively high startup delays, e.g. compared with TV watching. Downlink streaming of VR services uses this kind of approach too, often augmented with *Tiled Streaming* technology, which allows to divide the spatial content into pieces as well, called *tiles*, which too can be downloaded at different qualities, depending on

[1] https://www.forbes.com/sites/theyec/2019/07/08/extended-reality-xr-is-the-hot-topic-of-2020-and-beyond-heres-why/

[2] http://www.camerize.com/

[3] https://tools.ietf.org/html/rfc8216

[4] http://standards.iso.org/ittf/PubliclyAvailableStandards/c065274_ISO_IEC_23009-1_2014.zip

whether they are in the user's field of view[5]. Tiles which are not in the user's field of view are still downloaded (albeit at a lower quality), to avoid that an abrupt user head movement would result in the end user seeing a "black screen" even if only for a few (milli)seconds. While this approach addresses the quality of service delivered to the user, a certain amount of bandwidth is wasted sending over content (tiles) that most likely are not needed to be viewed.

For what concerns online uplink streaming services, the RTMP protocol[6] is often used. RTMP is optimized for low delays (i.e. realtime transmission) and does not handle bandwidth fluctuations very well. Since in current network it is not possible for a streaming client to select the best bitrate and resolution beforehand, the use of this protocol does not optimally employ the bandwidth that may be available and quality degradations (e.g. skipped video frames if bandwidth becomes too low) can hardly be prevented.

5G, and in particular network slicing, edge computing and orchestration, have the potential to provide service guarantees to media applications, which would alleviate – if not completely eliminate – the disadvantages described above. The release 16 of the 3GPP includes an initial version of a 5G Media Streaming Architecture[7], which describes some "application functions" and "application servers" that should be provided in media slices, both for downlink and uplink. However, fundamental questions such as how a media slice is set up and how to offer service guarantees over such a slice have not yet been answered in this standardization effort. Orchestration as such remains a key word: it becomes apparent that the orchestrator needs to deal with various "form factors": an advanced network service can be composed of appliances, bare-metal servers, virtual machines, stand-alone containers or containers bundles governed by their own orchestrator (think k8s). Complexity of such services can be high and the need for automation and orchestration only increases.

2.2 Existing test environments

In the context of 5G PPP Phase 3 projects, there are several H2020 projects that cater to provide end-to-end facilities to third parties for testing with 5G: 5G-EVE, 5G-VINNI and 5Genesis. These facilities focus especially on providing a baseline 5G infrastructure and providing access to vertical industry players to use them to test their vertical use cases. By contrast, the vertical testbed facilities offered within 5GINFIRE, and specifically the 5G Media Vertical testbed, focus on offering 5G environments already customized for specific vertical sectors, while at the same time abstracting several of the technological innovations of 5G in order to make it easier for vertical sector players to exploit the power of 5G and orchestration.

3 5G tests needed in Media

Tests in the Media vertical are needed in the following areas:

- Resource (bandwidth / latency / processing) guarantees: in order to deliver superior and new services with respect to what done today, it is imperative that the 5G network be able to provide certain, agreed service guarantees for media applications. For

[5] https://www.tiledmedia.com/index.php/technology/

[6] https://tools.ietf.org/html/rfc3550

[7] http://www.3gpp.org/ftp//Specs/archive/26_series/26.501/26501-g10.zip

example for professional-grade video production, bandwidth needs to be constant and high in order to guarantee the capture of studio-quality content; similarly, interactive VR streaming services need very low latencies in order to give the users the feeling of immersion; finally, other services (will) rely heavily on in-network processing capabilities, which need to be maintained throughout a service session for the service to be functioning correctly.

- Lack of interference among different (media) slices: once services have received a certain guarantee, no traffic or bottlenecks in other slices should interfere with the delivery of the services that are in different slice(s).
- Reactivity in the orchestration of resources (within a slice): the resources needed by a certain service may dynamically change during the delivery of the service (e.g. a camera feed may be upgraded to a higher streaming quality or downgraded to a lower one, depending on the importance of the content being captured live); the reaction time of the orchestrator needs to match the requirements for the service.
- Speed in the creation of a slice: tests should be conducted on verifying the time necessary to startup a slice; also, the tradeoffs between the speed of slice creation vs resources required to realize it should be analysed.

4 5GinFIRE current positioning in test environment

The 5G Media Vertical Testbed is located at TNO's research facilities and is connected to the federated 5GINFIRE infrastructure. TNO contribution provides 5GINFIRE infrastructure with a new, state of the art, testbed facility targeted at the 5G Media Vertical industry. This facility enables the execution of media use cases, specifically those requiring low latency, stable and high quality upload links. The testbed consists of a private cloud governed by the OpenStack Virtual Infrastructure Manager (VIM) which is connected to multiple physical eNBs via physical networks on one side and to the Internet on the other side. At the same time, for easier and faster prototyping, there is a possibility to use emulated eNBs along with emulated User Equipment (UE). The mobile network functions are part of TNOs Hi5 platform based on Fraunhofer's Open5GCore Release 3 which maps to 3GPP Release 14 and Release 4 will maps to 3GPP Release 15.

While the testbed is versatile and can in principle host various types of experiments related to the 5G networks, the unique selling point is a facilitation of the media-related workloads. Two advanced media-related services are provided:

- Live Uplink Streaming Service (LUSS)
- Virtual Reality Streaming Service (VRRS)

Live Uplink Streaming Service (LUSS) functionality is intended for experimenting with application- and network-adaptation in different uplink streaming scenarios. It assumes that several live streams are available (e.g., form a camera-crew members, which are regarded as clients of the 5G network) and one stream is picked to be broadcasted at a given moment. After one stream is selected to go live (i.e. stream in HQ), it could be that the current bandwidth is not sufficient for all the streams. Therefore, we introduce a bandwidth policing for the clients on the (emulated) Radio Access Network (which is usually the network's bandwidth bottleneck). Specifically, we police the uplinks of the UEs, assuring that the client with low priority has only limited resources in terms of bandwidth and cannot take them from the high priority client. At the same time, the clients which are not selected can be informed that they should decrease their encoding quality to save the bandwidth. Nevertheless, if for whatever reason the non-selected client tries to use more bandwidth (e.g., another application running on a device is trying to do the backup to the cloud), our enforced policy prevents any harmful effect of this.

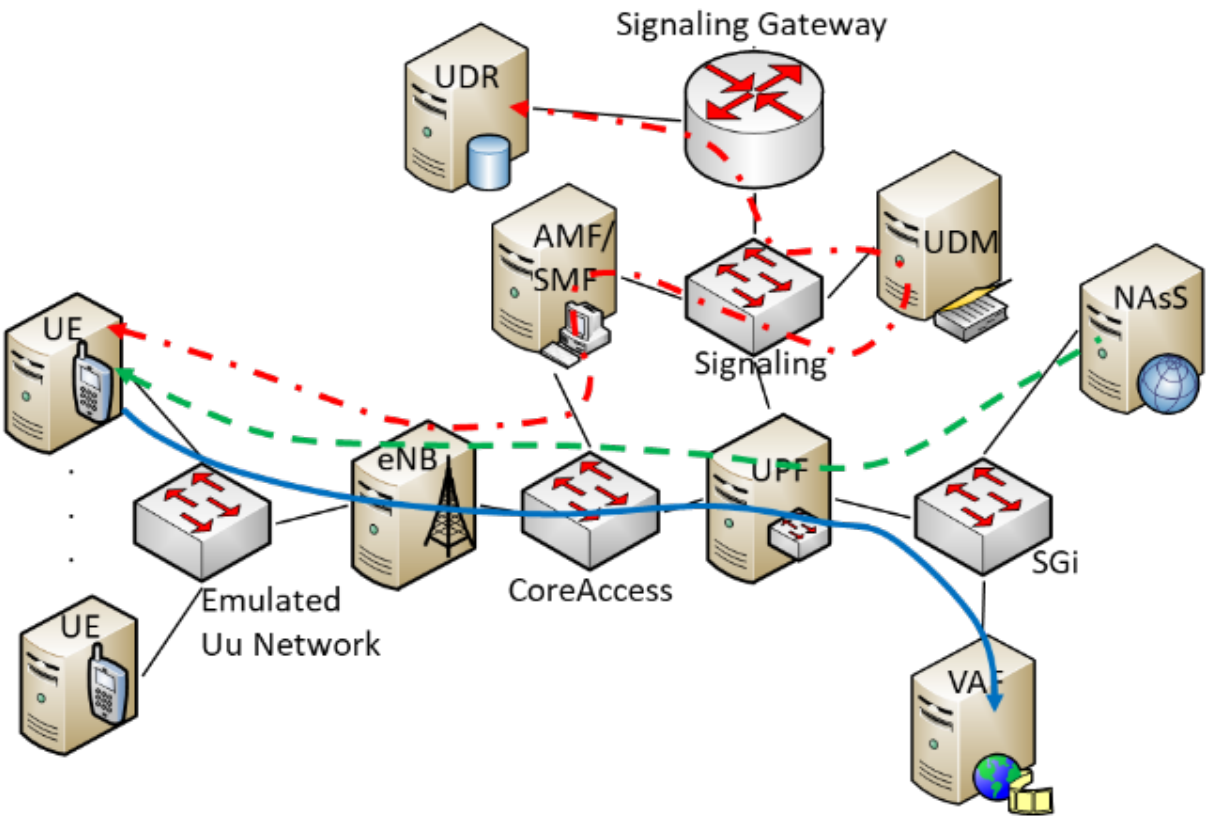

Figure 1: Live Uplink Streaming Service (LUSS) on top of 5G network service

Virtual Reality Streaming Service (VRSS). VR applications are known for their high bandwidth usage which can be reduced by streaming only specific region of interest (ROI) to the client, However, this approach creates an extra requirement in the form of latency because the time between the user changing his/her head position (i.e. changing field of view – FoV) and receiving the correct ROI must be as close to zero as possible to keep a good quality of experience. TNO's solution, the 5G Edge Masker (5GEM), streams a video which is still in an omnidirectional format, but where all regions except for the ROI are masked with opaque pixels, which enables the encoding algorithm to produce a content with a drastically reduced size. Furthermore, the Network Assistance Service (NASS) can request the network to change the path used to transport a video traffic to the low-delay one (which can be considered as a premium resource) if it detects that the client experiences too high delay.

5 5GinFIRE future developments

The 5G Media Vertical testbed is built on top of the platforms regarded as TNO's strategic assets for which every year a substantial investment budget is made available. The foreseen updates related to the Radio Access Network are increasing the number of the base stations. Some of them are expected to be placed in the main laboratory to test seamless handover and 5G session and service continuity features. Other base stations are planned to be then placed at other TNO locations thus extending geographical distribution of the testbed. We are also investigating possibilities to acquire Millimetre Waves antenna configuration in order to experiment with beamforming and massive MIMO. Cloud infrastructure is planned to be expanded with compute nodes featuring multi-terabyte RAM and GPU accelerators for both highly demanding media applications as well as AI/ML based algorithmic systems which can be used for management and orchestration. Storage system will be expanded with NVMe arrays to assure fast and reliable space for multi-terabyte media files (think point cloud, 360deg video). Key switching fabric elements are expected to be upgraded to 100Gbps to assure no bandwidth bottlenecks in the environment where many tenants require high network capacity. Finally, further investments in media-specific hardware such as VR/AR googles, haptic force feedback gloves or depth-cameras are budgeted.

6 Cooperation with other test environments

Jointly with several other partners (e.g., Economic Board Groningen, Vodafone, Ericsson), TNO is a 'founding father' of the 5Groningen[8] initiative which aims to provide a fieldlab to test applications running on top of the 5G networks. The applications concern developments in the media, care, energy, traffic & logistics, agriculture and living environment sectors. The region where a fieldlab is placed is sparsely populated rural area. This means is it complementary to the various available 5G testbeds located in the (densely populated) cities. Thanks to the involvement of the operators, vendors, system integrators, knowledge institutes and public parties, realistic, large scale tests can be scheduled in order to research the practical applications of 5G. The 5G Media Vertical testbed can easily be extended to 5Groningen fieldlab.

7 Tests services offer : current and future

The 5G Media Vertical Testbed extends the 5GINFIRE infrastructure with a new, state of the art, testbed facility targeted at the 5G Media Vertical industry. This testbed enables execution of advanced media use cases over a 5G network and an orchestrated cloud, with focus on low latency downlink streaming and stable and high quality uplink streaming. The exemplary services to be tested are related to:

1) 6-degrees-of-freedom (6DoF) VR streaming, where users can move in a virtual world (which is generated and rendered in the orchestrated cloud and transmitted with low latency to the user)
2) High bandwidth uplink streaming for professional video production

These kind of services one hand create opportunity for completely new user's experience, on the other hand pose a serious challenge to the infrastructure due to their very high and tight resource requirements.

To address these challenges, the design exploits programmability of the infrastructure to create the "app-aware" network slices which can adapt themselves to assure best Quality of Experience for the user. To assure ease of the deployment and reproducibility of the results,

[8] https://www.5groningen.nl/en ; for the full list of partners please visit https://www.5groningen.nl/partners

the services are instantiated via the orchestrator (OSM), following Infrastructure-as-a-Code principle. Experimentation and testing is facilitated by the (extensible) test suites based on the ROBOT framework.

The 5G Media Vertical testbed is designed in such a way that, next to the use cases presented, other use cases may also be implemented by the experimenters.

The team behind the 5G Media Vertical Testbed has knowledge in both media and infrastructure domains and experience in bringing these two fields together. The experimenters can enjoy various benefits from using the 5G Media Vertical Testbed: those who lack expertise in 5G and orchestrated cloud technologies can simply relay on the underlying infrastructure and testbed experts while focusing on the media layer. Also, easy-to-use templates and APIs facilitating 5G slices deployments are available in the testbed, which allow for a certain level of 'self-service' with respect to 5G infrastructure, while shielding the complexity of the network.

8 Feedback from market

From TNO's experience, while there is a huge potential in 5G networks for verticals, many vertical actors do not understand the details of the underlying mobile network technology, and in fact they are not interested in understanding them. Instead, the verticals would like to be able to express their services descriptions and requirements in the language that is understood to them (think: number of cameras, video quality, location of the event) and off-load a translation to a 5G-network-service language to some other entity. We see two ways to address this problem: on a short term, a team of qualified researchers from TNO can facilitate onboarding and testing the functions delivered by a vertical party. On a mid-term, TNO is involved in developments of the "5G-as-a-Service" concepts where an intermediate platform exposes "vertical-friendly" north-bound APIs to make the definition of the services easy. Under the hood, the platform is to convert them into the requests understood by the orchestrator, composing a service consisting of 5G network building blocks and taking constraints such as bandwidth, delay budget or physical location into consideration.