Preliminary Evaluation of a Software-based 5GC Release 17 5MBS Prototype

Carlos Barjau[†], Borja Iñesta[†], David Gómez Barquero[†], Pousali Chakraborty*, Hemant Zope*, Marius Corici *

[†]iTEAM Research Insitute,Universitat Politecnica de Valencia
Valencia, Spain
Email: {carbare1,dagobar,borieher} @iteam.upv.es

* Fraunhofer FOKUS Institute
Berlin, Germany
Email:{pousali.chakraborty, hemant.zope, marius-iulian.corici}
@fokus.fraunhofer.de

Abstract— **Point-to-multipoint communications were introduced in the 5G System as part of Release 17. This technology is known as 5G Multicast-Broadcast Services (5MBS) and provides a scalable procedure to reuse New Radio in mass delivery of multimedia. In order to evaluate the impact of Point-to-multipoint communications in the 5G Core, a 5MBS software prototype was implemented by extending existing Point-to-point Network Functions of Open5GCore, a commercial 5G Core solution. A performance comparison between Point-to-point and developed Point-to-Multipoint components has been carried out; using network traffic tools and video delivery. The results show that the 5MBS prototype can course traffic to multiple users simultaneously without demanding more network bandwidth on different topologies, but the maximum bitrate coursed is degraded when compare versus Point-to-point components.**

Keywords— 5G, 3GPP, Release 17, Broadcast, Multicast, 5MBS

I. INTRODUCTION

Nowadays, cellular traffic is mostly made of multimedia video with an ever-increasing demand year to year; not only affected by the quantity of videos watched but the higher quality formats such as 4K, or VR [1]. This video demand can be separated into on-demand content, where the content is stored and delivered based on user requests; or live, where users tune into a continuous transmission e.g. Livestreaming. To cope with this, mobile operators have been buying more spectrum and invested into further spectral efficient radio techniques e.g Massive MIMO and Beamforming. But there is another way to provide an scalable mechanism for live multimedia content: Point-to-multipoint communications, where the content is only delivered once over the radio and core network towards the users consuming the content.

3G and 4G already included this possibility, known as Multicast-Broadcast Multimedia Services (MBMS) [2] and its evolution enhanced MBMS (eMBMS) [3]. These solutions targeted Digital Terrestrial Television (DTT) services as their main goal. eMBMS standard defines the optional extensions to the baseline cellular architecture, introducing new elements both at radio and core while adapting the air interface towards broadcaster infrastructure usually formed by High Power High Tower deployments. The situation in 5G is different, where initial versions of the cellular standard did not define a Point-to-multipoint framework. 3GPP argued that the latest versions of eMBMS, known as Enhanced Television (ENTV) [4] where mature enough to meet 5G requirements. Nevertheless, a study item in 3GPP was created to evaluate the inclusion of Point-to-multipoint in 5G, which derived in a suite of 5G-native specifications called 5G Multicast-Broadcast Services or 5MBS [5]. In parallel, eMBMS keeps receiving updates and it is now called 5G Broadcast as the technology has matured enough to be the DTT technology to provide nation-wide public linear TV.

This paper evaluates the impact of introducing multicast in the 5G Core (5GC) in order to model the performance gain and bandwidth savings by implementing a software based 5MBS prototype. It has been structured as follows: Section II describes how 5MBS has been defined and Section III detail the software implementation procedure followed and initial results comparing the Point-to-multipoint components against the Point-to-point ones.

II. 5MBS STATE-OF-THE-ART

5MBS was introduced in Release 17 as a way to provide Point-to-Multipoint communications in the 5GS [6]. The scope of the technology is to provide a dynamic transmission mode between unicast, multicast and broadcast to adapt for several services while reusing as much as possible from the 5G System. In this sense, covering the radio, core and service with enhancements. The technology consists of several additional functionalities; extensions to existing Network Functions (NFs) and new ones, logical communication modes, types of transport, sessions, delivery modes, protocols, interfaces, and three different type of RAN communications. To summarize, three communication modes are defined: Broadcast, multicast and unicast; a new type of session, named MBS session, is

979-8-3503-2152-4/23/$31.00

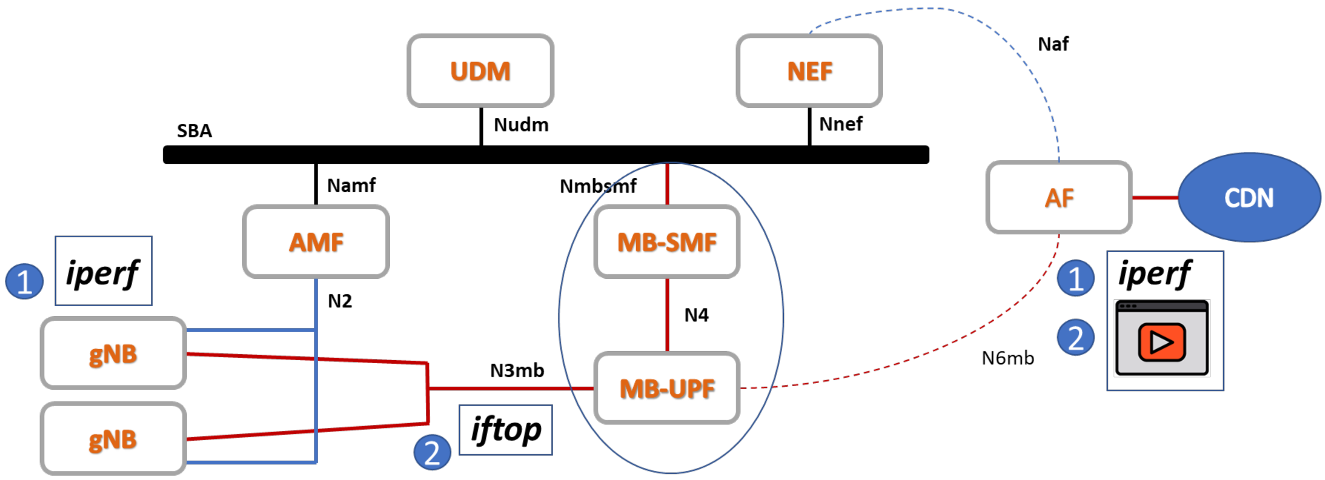

Figure 1: 5MBS prototype programmed and implemented with associated interfaces. The prototype has been implemented by extending commercial 5G Core product Open5GCore, in particular, the Multicast/Broadcast Session Management Function (MB-SMF) and the Multicast/Broadcast User Plane Function (MB-UPF). The two validation experiments performed are the latency and maximum bandwidth (1) and a video delivery test to several virtual users (2).

introduced to manage outgoing 5MBS services. Two delivery modes in the 5GC have been standardized, Shared Delivery where 1 copy of a packet is delivered to multiple gNBs and Individual Delivery where 1 copy of a packet is delivered to one or more legacy User Plane Function (UPF) which will deliver the data to a legacy NG-RAN that does not support 5MBS.

Inside the 5GC, 5MBS defines four new NFs: the Multicast-Broadcast UPF (MB-UPF), the Multicast-Broadcast Session Management Function (MB-SMF), the Multicast-Broadcast Session Function (MB-SF) and the Multicast-Broadcast Session Transport Function (MB-STF). MB-UPF and MB-SMF can be considered part of the transport layer side of the 5GC, while MB-SF and MB-STF compose the service layer aspect of the 5GC. The changes inside the 5GC is not only limited to new NFs, but also include new functionalities to existing NFs such as the Access and Mobility Management Function (AMF), Policy Control Function (PCF) and the Network Exposure Function (NEF). Additionally, new interfaces connecting the 5MBS specific NFs and existing ones have been defined. Regarding the NG-RAN and Physical Layer, the most relevant change is that New Radio waveforms and procedures have been reused in order to ease manufacturer adoption; while in the gNB some new procedures have been defined to ensure lossless handover, retransmission schemes and seamless switching.

III. 5MBS PROTOTYPE AND LATENCY PERFORMANCE

In order to evaluate the impact of Point-to-multipoint communications in the 5GC, a software prototype of the 5MBS transport layer functions i.e. the MB-UPF and MB-SMF were programmed. It was chosen to extend existing NFs from a commercial 5GC product, the Open5GCore [7] by Fraunhofer Fokus, which implements a 3GPP compliant Release 16 Core. The functions extended were FOKUS UPF and SMF, alongside modifications to the AMF and NEF to support the MBS sessions. The NG-RAN and the device side are emulated and also provided by Open5GCore, known as the Benchmarking Tool. The virtual gNB and devices have been modified to accept multicast packets from the Application Function (AF). Moreover, N3mb now supports the delivery of multicast packets from the MB-UPF to the virtual gNBs. A diagram of the prototype can be seen in Figure 1.

Table 1: 5MBS iperf results using UDP traffic over unicast and multicast.

Type of Traffic	Bandwidth	% of lost datagrams
Unicast	1 Mbit/s	0%
	2 Mbit/s	0%
	5 Mbit/s	0%
	10 Mbit/s	0%
	15 Mbit/s	0%
	20 Mbit/s	0%
	35 Mbit/s	0.88%
	50 Mbit/s	14%
Multicast	1 Mbit/s	0%
	2 Mbit/s	0%
	5 Mbit/s	0%
	10 Mbit/s	0%
	15 Mbit/s	0%
	20 Mbit/s	0%
	35 Mbit/s	0.75%
	50 Mbit/s	22%

The preliminary test performed was an *iperf* server test between the AF connected to the MB-UPF over the N6mb interface, to a virtual device connected to a gNB hosting the *iperf* client. In the test, it was evaluated the impact of using the baseline FOKUS SMF and UPF against the modified FOKUS MB-SMF and MB-UPF. The prototype is running in a server-grade rack, which has Proxmox as its hypervisor and provides a Virtual Machine with Ubuntu 16.0 as its operating system. This VM has allocated 4 processors, 2.1 GHz and 4 GB RAM. Every function and virtualized RAN is running locally but

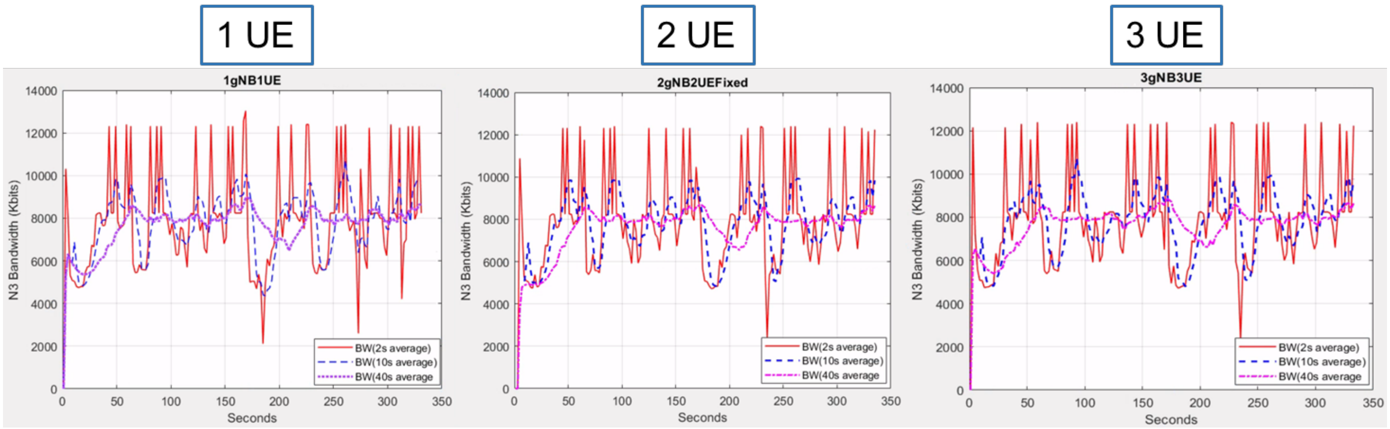

Figure 2: Bandwidth measurements over the N3mb interface of the 5MBS Open5GCore prototype, with a variable number of virtualized UEs. From left to right, 1, 2 and 3 UEs have been attached to 1, 2 and 3 gNBs respectively. Linux software iftop has been used to gather the BW measurements, with the red straight line representing the 2 seconds average, the blue dashed line representing the 10 seconds average and the magenta dotted line representing the 40 seconds average.

abstracted from each other using network namespaces. *iperf* Linux command has been run with different bandwidth values, using UDP traffic for every test. Each test has been executed for 300 s. The results can be seen in Table 1. It can be seen that for the ranges from 1 to 20 Mbit/s there are no lost datagrams, however as the bandwidth reaches 35 Mbit/s some UDP datagrams start to get lost. When the bandwidth is increased to 50 Mbit/s, the performance [8] degradation of multicast is apparent, reaching 57% more datagram loss compared against unicast.

IV. VIDEO DELIVERY PERFORMANCE

To perform the validation of the prototype described in Section III, two different video delivery tests were carried out, with different RAN-UE topologies to evaluate diverse scenarios. For both tests, the AF in Figure 1 acts as a video server, using *ffmpeg* to encapsulate and send the video to the MB-UPF via the N6 interface. The MB-UPF will receive this video, recognize it coming from a Multicast IP, then forward it to the virtualized RAN components. The UEs receive the video stream using *ffplay*. Depending on the topology used in the virtualized RAN, two tests have been devised and described below.

The first one, where a single service is delivered to a variable number of virtualized UEs, which are attached to a variable number of gNBs following a 1:1 mapping i.e. UE1 attach to gNB1, UE2 attach to gNB2... In detail, the video used in the experiments is a mobile video recorded at 4K quality, with 2:47 minutes duration, H.264 as compression codec and 6 Mb/s average bitrate. The video is run twice, and started manually, to see the effect of the bandwidth drop over the N3mb interface. A variable number of virtual UEs have been tested: 1, 2 and 3. Figure 2 shows the bandwidth captured for each experiment.

The second test is composed of 2 UEs consuming 2 distinct services i.e. the same video sent over two different multicast IPs, but the number of gNBs is variable between 1 and 2. In this case, the goal was to evaluate if the number of gNBs connected over the N3mb interface does affect the bandwidth coursed. The video chosen was a lower bitrate compared to previous validation test in order to avoid the transmission peaks, with an average bitrate of 1.6 Mb/s and in a similar way, it is launched twice for each service to double the duration. The bandwidth captured during the experiment can be seen in Figure 3.

From Figure 2 graphs, three main points can be derived: First, the decrease in bandwidth in the middle of graphs corresponds to where the video is relaunched manually. Depending on the timing and the processor queue, the stop and correspondent drop in bandwidth vary as well. Second, it can be seen that the transmission peaks around 12 Mbit/s. This value may be the maximum bitrate that be coursed with the current 5GC parametrization, including the VM capabilities and the abstraction layers added. Further studying of different video qualities, VM characteristics or other deployment topologies may influence the maximum bitrate value. Last, while the 2 seconds average line from graph to graph slightly varies from case to case, the longer averages highlight how the bandwidth coursed is independent of the UEs which have joined the 5MBS transmission. This provides validation that the MB-UPF is not replicating packets per UE and it is using a single tunnel from the Core to the RAN to deliver the video. It can be concluded that multicast saves bandwidth resources in the data plane when users consume live content.

Regarding Figure 3 graphs, while the peak values coincide, the shape between the “2UEs 1gNB” and “2UE 2gNB” is fundamentally different. The reason for this is the human factor, where the tests have not been automated and for each video service, the virtualized UEs, monitoring tool launch, stop and restart needs to be performed manually, introducing randomness into the measurement experiment. Nevertheless, the graphs show both bandwidth reductions around the middle duration, marking the restart time of the videos inside the prototype. The main conclusion derived from this experiment is that the number of gNBs attached to the MB-UPF does not affect the bandwidth coursed over the N3mb interface.

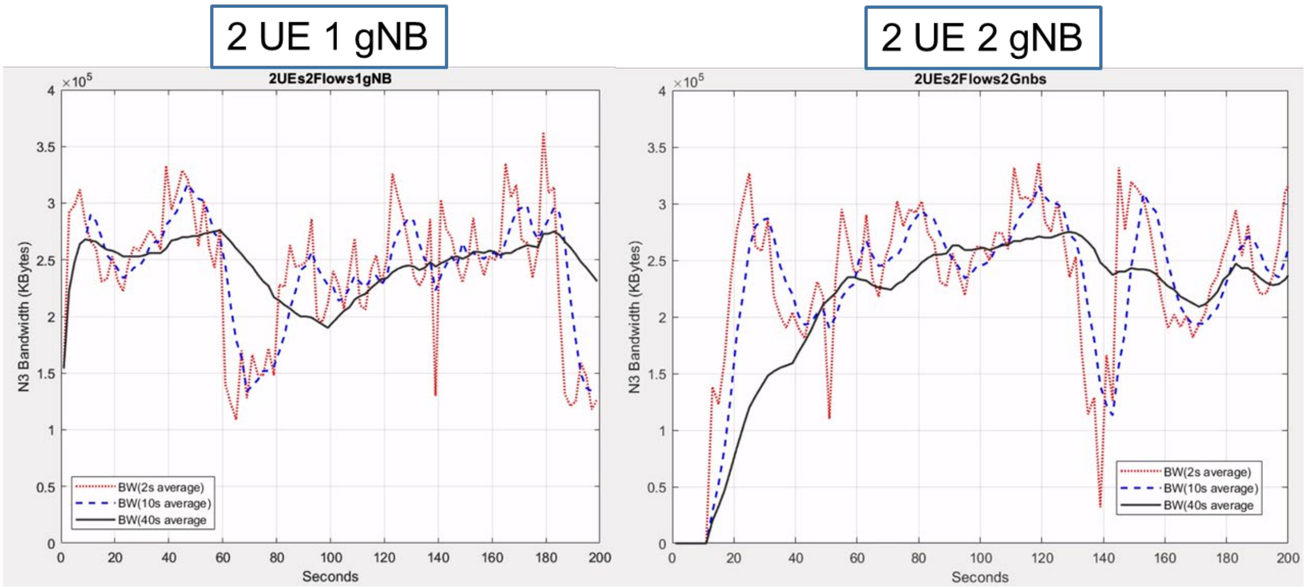

Figure 3: Bandwidth measurements over the N3mb interface of the 5MBS Open5GCore prototype, with 2 UEs, a variable number of virtualized gNBs and two multicast services. From left to right, 1 and 2 gNBs have been attached to the MB-UPF over the N3mb interface. Linux software iftop has been used to gather the BW measurements, with the red straight line representing the 2 seconds average, the blue dashed line representing the 10 seconds average and the black straight line representing the 40 seconds average. Both multicast services are relaunched around the 140 second mark and left running until the second 200.

V. CONCLUSION

In this paper, the implementation and evaluation of a 5MBS prototype has been presented. 5MBS was introduced during Release 17 so no commercial equipment was available to evaluate the impact of point-to-multipoint in the 5GC. A software-based solution was made based on an existing commercial 5GC, Open5GCore. 5MBS basic concepts are outlined in section II. Section III describes the prototype and a evaluation comparison between the unicast counterpart. In the full version of the paper, further results including video delivery over the N3mb interface will be included alongside appropriate references.

VI. ACKNOWLEDGEMENTS

This work was partially supported by the European Commission under the 5G-PPP project FUDGE-5G (H2020-ICT-42-2020 call, grant 957242) and the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU through the UNICO-5G I+D AVANZANDO-5G-INMERSIVO project (TSI-063000-2021-109, TSI-063000-2021-110, TSI-063000-2021-111).

VII. REFERENCES

[1] Ericsson, "Mobile data traffic outlook," 2022. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast.

[2] F. Hartung, U. Horn, J. Huschke, Kampmann and T. Lohmar, "MBMS - IP Multicast/Broadcast in 3G Networks.," *International Journal of Digital Multimedia Broadcasting* , vol. 1, 2009.

[3] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements," *IEEE Communications Magazine,* vol. 50, no. 11, pp. 68-74, 2012.

[4] D. Gomez-Barquero, J. J. Gimenez and R. Beutler, "3GPP Enhancements for Television Services: LTE-Based 5G Terrestrial Broadcast," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, Ltd, 2020, pp. 1-10.

[5] 5G-MAG, «5G-MAG Explainer: Media distribution with 5G Multicast-Broadcast Services,» [En línea]. Available: https://drive.google.com/file/d/1OJkHzjXeI9SrlXE98D7kxdD7b_m6dxXS/view.

[6] A. Rico, I. Bouazizi, M. Griot, P. Kadiri, L. Liu y T. Stockhammer, «3GPP Rel-17 Extensions for 5G Media Delivery,» *IEEE Transactions on Broadcasting,* vol. 68, nº 2, pp. 422-438, 2022.

[7] Fraunhofer FOKUS, «Open5GCore - 5G Core Network for Research, Testbeds and Trials,» 2022. [En línea]. Available: https://www.open5gcore.org/.

[8] Advanced Television, " ZTE claims 5G NR broadcast first," 2020. [Online]. Available: https://advanced-television.com/2020/09/30/zte-claims-5g-nr-broadcast-first/.