On Minimizing Service Access Latency: Employing MEC on the Fronthaul of Heterogeneous 5G Architectures

Nikos Makris[†⋆], Virgilios Passas[†], Christos Nanis[†], Thanasis Korakis[†⋆]
[†]Department of Electrical and Computer Engineering, University of Thessaly, Greece
[⋆]Centre for Research and Technology Hellas, CERTH, Greece
Email: {nimakris, vipassas, cnanis, korakis}@uth.gr

***Abstract*—Multiple-access Edge Computing (MEC) has been proposed as a means to minimize the user to service path latency, by deploying and operating datacenter resources close to the network edge. The introduction of 5G mobile network services, and their provisioning through disaggregated base stations complying with the Cloud-RAN paradigm, allows the redefinition of the traditional Edge Computing by offering deployment of services even closer to the network access edge. In this work, we leverage a disaggregated heterogeneous 5G infrastructure, compliant with the 5G New Radio (NR) specifications, and present a scheme for placing the services even closer to the Edge, close to the concept of fog computing. We develop a scheme for the OpenAirInterface platform that allows services to be executed close or over the machines hosting the radio services for the network access. By exploiting features for integrating heterogeneous radio resources in the cell, we are able to create a controller interface for selecting the optimal radio access technology used to serve each user of the network from the MEC service perspective. We evaluate our solution in a real testbed setup, and measure performance related indicators for our solution by using adaptive video streaming. Our results illustrate up to 80% better video qualities delivered to the end user when appropriately selecting the access technology.**

***Index Terms*—Multi-access Edge Computing, Cloud-RAN, low latency, 5G, OpenAirInterface**

I. Introduction

5G Mobile Networks are expected to bring several advancements towards providing higher network speeds with lower latency over the network. These aspects will allow time critical applications to run over this system, bringing further innovation for application providers and vertical services, hosted over this infrastructure. Low-latency requirements is expected to be fulfilled through the wide proliferation of edge computing; services being deployed closer to the network edge, will be able to serve users with lower response times and with content that is appropriately replicated at the edge datacenters. User access will consist of several different technologies for wireless, such as the forthcoming 5G New radio (5G-NR) or the legacy LTE and WiFi technologies. As a matter of fact, ETSI has revised the annotation for Mobile Edge Computing to Multiple-access Edge Computing (MEC) [1] in order to reflect on the different technologies used for user access.

At the same time, 5G brings new architectures for the operation of base stations, adding up to the flexibility and management of the distributed infrastructure. In this paper we deal with the disaggregation of base station unit based on the Cloud-RAN concept [2], allowing the operation of lower intelligence units at the network edge, and moving the processing tasks to the Cloud. The disaggregation of base stations allows the re-conception of technologies such as MEC, towards placing the provided services deeper in the network. The specifications for the new 5G radio access protocol (5G-NR) detail the architecture and interfaces for the intercommunication of the different base station entities; they define the disaggregation at the higher Layer-2 of the OSI stack, realizing the 3GPP suggested Option-2 split [3], between the Packet Data Convergence Protocol (PDCP) and Radio Link Control (RLC) layers of the mobile stack. The output of the splitting process are the Central Unit (CU) integrating the PDCP and above layers, and the Distributed Unit (DU) that handles the lower transmission and reception functions, up to the RLC layer. The specifications provide hooks for heterogeneous DUs to be integrated, such as 5G-NR, LTE and WiFi, managed by a single CU and Core Network.

Although MEC is considered as a low-latency solution in the drafting of the 5G-NR specification, its integration to the architecture is similar as with the legacy technologies. For example, the solution that is suggested for moving the MEC services closer to the base station is the *bump-in-the-wire* approach [4]: this approach considers intercepting the traffic between the base station and the core network (S1AP traffic) and redirecting it to the MEC service. For the case of a Cloud-RAN setup, ETSI has tailored the recommendation to deploy the MEC service together with the CU at the edge datacenter, intercepting again S1AP traffic. This creates a path UE-DU-CU-MEC in order to access the service, whereas the disaggregation of the base station could be leveraged to place the services deeper in the network, towards providing a path UE-DU-MEC. In this work, we experiment with placing the services on the fronthaul interface of a CU - heterogeneous DUs setup, extending the initial prototype presented in [5]. The platform that we use is OpenAirInterface [6], that provides an open source implementation of the base station stack. This paper contributes with additions to this prototype in order to:

- Include new signaling for supporting multiple applications and users
- Couple it with a multi-technology base station allowing seamless switch of technologies supporting the end users
- Provide a solution for switching the technology serving each end user for minimizing UE-to-service and vice-versa latency.
- Illustrate how wireless technology selection can widely affect user perceived quality, when accessing MEC resources.

II. Motivation and Related Work

Integrating edge computing resources, closer to the access network has been gaining more attention as the technology providers and network operators desire to minimize the UE to service latency. Works [7] and [8] highlight this fact, with contributions on the definition of interconnection interfaces with the 5G network. The different locations for deploying the MEC services have been discussed in [1] and [4], providing some use case specific performance indicators for several

5G applications (e.g. Industry 4.0, eHealth, AR/VR, etc.). These deployments are summarized in the following: 1) the *bump-in-the-wire* method, where the MEC service is placed on the backhaul link of the base station, interconnecting it with the Core Network and intercepting the S1-U traffic of the cellular network, 2) collocating the core network and the MEC servers at an Edge datacenter, and 3) using a distributed core network and control the MEC service through the PDN GW. Virtualization adds up to the flexibility of the network topology, allowing live of services to datacenters located closer to the UEs, based on their trajectory, etc.

Cloud-RANs redefine the base station architecture, and allow the splitted elements to be placed at different locations. Although different splits have been proposed [9], the split between the PDCP and RLC layers is standardized in the 5G New Radio (NR) specifications [10]. The specifications provide also hooks for non-3GPP technologies to be integrated as new DUs, similar to the LTE WLAN Aggregation Adaptation Protocol (LWAAP) for legacy LTE [11]. Despite the base station disaggregation, possible MEC deployments do not consider moving the edge services closer to the DU; in the best case, the services are co-located with the CU at the edge datacenter. The F1-U traffic carrying the user plane data is using GPRS Tunneling Protocol (GTP) tunnels, similar to the S1AP protocol that transfers data from the base station to the Core Network. Thus, engineering a solution for bringing the services at the true edge of the network and closer to the DU should not pose a big overhead for technology providers.

In this paper, we propose moving the MEC services closer to the true network edge, extending the prototype built in [5] and [12]. In [12], we showcased an implementation that integrates non-3GPP technologies (WiFi) to the disaggregated base station. The CU side is managing both DUs, and can steer traffic to the DU that will serve the network's UE on a per-packet basis. This prototype provided proof-of-concept experiments determining the maximum distance between the CU and heterogeneous DUs so as no service disruption is experienced at the UE side. This implementation used TCP/IP channels for the data plane communication between the PDCP and RLC layers of the stack, providing dedicated signaling for this purpose. This signaling is referred as **F1 over IP (F1oIP)**, as it has a similar structure with the standardized F1AP [10]. Since these splits use Ethernet based encapsulation, they can be easily handled by services introduced in the fronthaul interface. In [5], a prototype implementation was introduced to offer MEC services closer to the DU side of the network. Some indicative results showed that even for 10MHz channel bandwidth in LTE implementations of the DU, the UE to MEC service latency can drop below 10ms, sufficient to serve several 5G applications in terms of latency [4]. However, the solution does not integrate non-3GPP solutions in the RAN. Similar works on the development of similar MEC functionality in such experimental setups include [13] and [14]; in the former, the authors employ SDN based assisted control of GTP packets inside the Core Network, and in the latter the authors implement the "bump-in-the-wire" method to intercept packets on the backhaul interface of an LTE eNodeB. On the other side, in [14] the authors place the MEC services between the Core Network and the eNodeB. This allows to bring the service even further to the network Edge, however, the solution relies only on application space based management of the GTP tunnels that raises several performance issues, such as limited UE-to-service throughput and higher latency. In this work, we present a prototype that allows placing MEC functions on the fronthaul interface of heterogeneous disaggregated base stations, as shown in Figure 1. Multi-homed UEs can be served by either the cellular or the WiFi DUs, towards enhancing the UE to service path latency.

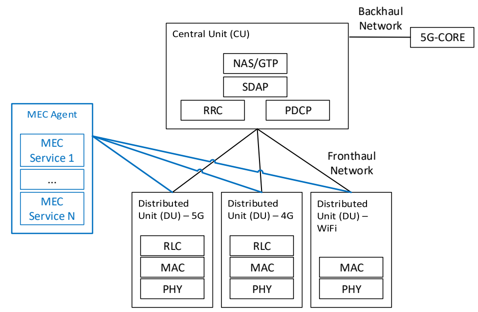


Fig. 1: Proposed base station architecture with MEC services being placed on the fronthaul interface

III. MEC System Architecture

The system architecture that allows us to place the services on the fronthaul interface is summarized in Figure 2. It consists of different elements orchestrating the CU - DU intercommunication, a MEC Agent handling the delivery of user data to services running on the edge and communicating with the DUs, and a mapping system to address network UEs based on their low-level L2 information (MAC address for WiFi, RNTIs for cellular network). For the implementation we employ the OpenAirInterface [6] platform, that is providing a software based implementation of the LTE networking stack. In the sections below we detail each element needed.

A. CU - DU communication

Based on the specifications of 5G-NR, the base station should be able to incorporate different technologies through the control of new DUs from the same CU. As the proposed split option between the PDCP and RLC stack has slack requirements for the fronthaul link, it is an excellent candidate for accommodating multiple technologies, even non-3GPP compliant, like WiFi. In fact, the legacy LTE protocol is using the PDCP layer as the convergence layer for integrating WiFi in the RAN [11]. In the overall system communication between CUs and DUs, the relationship 1:n, meaning that multiple DUs can be connected to a single CU. From the DU's perspective, this relationship is 1:1, so that each DU is associated only with a single CU.

In [12] we provided the F1oIP protocol as a communication mechanism between the CU and DUs. The software is handling the Service Access Points (SAP) between the PDCP and RLC layers: these are the *pdcp_rlc_data_request* for the Downlink (DL) traffic, and the *rlc_pdcp_data_indication* for the Uplink (UL) case. Instead of the default SAPs, we introduced a communication mechanism based on asynchronous sockets between the two layers. Such a mechanism allows

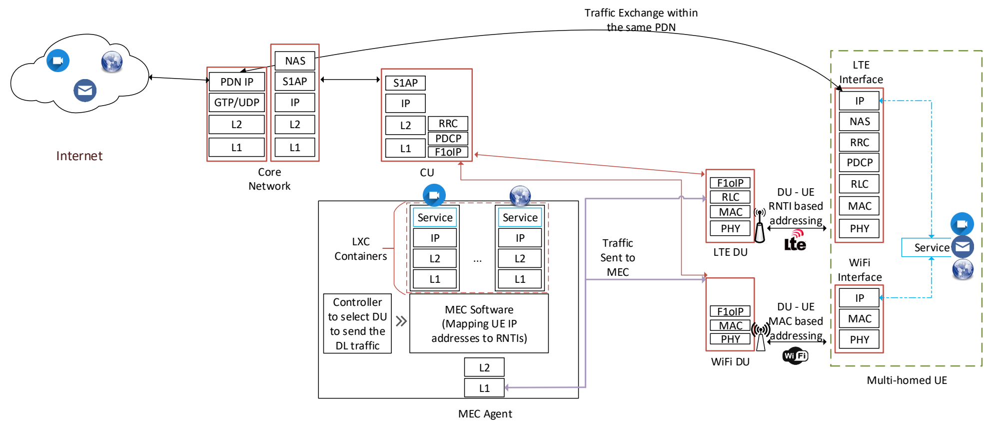


Fig. 2: Overall system architecture: F1oIP orchestrates the intercommunication between the CU and DUs, whereas the MEC software is providing the functionality of mapping RNTI to IPs

us to integrate other technologies by using an IP interface at the PDCP side. Under a monolithic setup, using the SAP interfaces, scheduling information is exchanged between the two layers intended for mapping the traffic to the logical, transport and subsequently physical channels of the network. The F1oIP implementation is piggy-backing this information in order to make such a transmission possible. The payload of these packets is a PDCP encapsulated packet, bearing a 2 byte long PDCP header. For the case where the CU is managing a non-3GPP DU (e.g. a WiFi DU), the same information is transmitted to the DU, but stripped off before injecting the traffic to the WiFi network. This information is further used in order to orchestrate the proper operation of the UL, by forming packets piggy-backing the information expected at the CU side. For the UL case, the reverse process takes place before transmitting the packet to the CU. For the WiFi integration case, the WiFi DU software generates new PDCP numbers, based on the traffic flow, generates the PDCP header and piggy-backs the information on the packets sent to the CU.

B. DU-MEC communication

As we focus on placing user accessible services on the fronthaul interface, we need to develop the respective functionality between the DUs and the MEC server hosting the provided services. In [5] we developed a similar protocol for the DU to MEC communication, by introducing a *MEC Agent* component. The agent can generate and exchange the appropriate messages with the DUs, or receive and deliver the respective data packets to the hosted MEC services.

The communication between the MEC agent and the DUs is similar as with the CU-DU case; when a DU has data to transmit to the MEC service creates a *mec_data_request* message. This message is then handled by the MEC agent and its payload (user data packets) is delivered to the service. Similarly, for the reverse path, the MEC agent generates a *mec_data_indication* for the DU that the client is registered with. The DU information is dynamically discovered, based on the DU address that the agent received a message from.

An important aspect here to consider is the ciphering process taking place at the PDCP layer of the network. As user plane data passes through the PDCP entity of the base station or the UE, it is ciphered according to the EPS Encryption Algorithm (EEA) chosen. Typically, there are four different variations of algorithms that are used (EEA0 - Null Ciphering Algorithm, EEA1 - SNOW 3G, EEA2 - AES, EEA3 - ZUC). Therefore, the data that is exchanged below the PDCP point are ciphered, and the proper decryption mechanisms need to take place in order to retrieve the user data. For this purpose, we introduce a control packet that is broadcasted from the PDCP entity to all the DUs and MEC agents that are operating in the system, in order to ensure the deciphering process. As this process is introducing extra delays for our experimental setup, we do not employ any PDCP encryption mechanism in the results presented in the current paper. Similar to the encryption case, we introduce extra signalling across the different entities of the network (CU, DUs, MEC Agents) in order to accommodate multiple clients over heterogeneous DUs. This includes mapping a cellular network UE with its respective non-3GPP interface and the manner that the different DUs identify it. We further detail how this is achieved in the following subsection.

C. Support for Multiple multi-homed UEs

Our target setup is considering multi-homed UEs, with network service over more than one radio access technology concurrently. Cellular base stations are merely seen as a Layer 2 device from the UE side: the end-to-end connection established between the UEs is with the core network in the context of a PDN (see Figure 2). Each PDN is a separate broadcast domain, and all entities under it (Core Network and UEs) can communicate with each other. Therefore, and as the MEC agent interfaces only DUs from the RAN side, the data coming from the cellular RAN is only interfaced through cellular network L2 information, i.e. the Radio Network Temporary Identifier (RNTI). This RNTI is used by the base station for forwarding the user plane data to the UEs, mapping them to the different logical and transport channels, etc. Contrary to this, for the WiFi case and the MEC services, the UEs are identified using IP addresses. This allows them to be solely addressed and request services from the MEC agent, based on the IP configuration of the service.

In order to cope with this problem, we introduce new signaling to the network as follows: whenever a new client registers with the cellular DU and a new RNTI is allocated, we transmit an *rnti_inform* message to all the DUs and MEC agents. The message contains the RNTI information, a UE id based on the sequence of attached UEs and the DU with which it is associated to. With this information we create a mapping between the RNTI and the IP address that will be allocated by the Core Network to the UE, and be able to distinguish between them during the operation of the MEC agent. The RNTI information is actually being piggy-backed by both the DUs and MEC Agents of the network when sending data to the CU or the cellular DU respectively. Through this mapping we can use multiple services offered to multiple UEs, connected with multiple technologies. This functionality makes use of a separate control channel introduced between the CU, DUs and MEC agent of the architecture. This allows us to expose an API at the MEC Agent level to select the technology through which each user will be served in the wireless domain for the DL MEC traffic. By sending to the agent a specific UE identifier and the selected technology that it will use, the MEC agent updates its mapping for the specific UE and in case of traffic being sent for the DL communication path, the data is sent to the DU with the technology denoted by the controller. In case that there is no such selection, the agent replies through the DU via which initially the UE transmitted traffic.

D. Support for multiple MEC services

The MEC agent software detailed above is one of the key software components facilitating hosting the services on the fronthaul network. Whenever the agent receives traffic intended for the hosted services, it decapsulates it and injects the user payload to the MEC service. We select to host the services containerized through Linux Containers (LXC), as they can be instantiated on the fly, whenever an end-user requests different services from the MEC platform. Adopting LXC containers is very beneficial as it allows each new service to be addressed with a new container, with a new network IP address and can be easily migrated if needed to another edge host, like for example in the case of a rapidly moving mobile UE (V2X case). As the LXC service places all the containers with different IP addresses under a bridge interface on the edge host, the MEC agent has to inject the traffic to the bridge, destined to the MAC address of the container hosting the MEC service. Through the RNTI - IP address mapping described in the previous section, multiple UEs can make use of the same service, even when they are getting connectivity through different access technologies.

IV. Experimental Setup

In this section, we present our experimental setup and experimentation methodology. The functionality has been developed over the OpenAirInterface platform (OAI) [6], that provides an open source software implementation of the cellular base station stack and can be executed over commodity hardware with the appropriate Software Defined Radio front-ends. We conduct the experiments over the NITOS testbed [15]. NITOS is a heterogeneous testbed located in University of Thessaly in Greece, offering a rich remotely accessible experimentation environment with resources spanning from commercial LTE, to WiFi and Software Defined Radio platforms.

We focus on the LTE implementation of OAI, as it provides the functionality for the high layer splits compared to the recent 5G-NR release. We employ an altered version of the WiFi DU developed in [12] in order to setup a separate communication channel between each DU and the MEC Agent, and a control channel between the CU and all the DUs that transmits the RNTI related information for UE to service mappings. This channel, and the F1oIP channels for the CU/DU communications are selected to be TCP over Ethernet, as our former experiments in [12] denote that there is no notable performance degradation compared to UDP or even the vanilla OAI setup.

TABLE I: Equipment parameters

Network Parameters	Values
LTE mode	FDD Band 7
LTE Frequency	2680 MHz (DL)
LTE Antenna Mode	SISO
Number of RBs	50 (10 MHz)
UE	Cat. 4 LTE, Huawei E3272
WiFi Technology	802.11n MIMO 3x3
WiFi Channel BW	40 MHz
WiFi card	Atheros 9380
Backhaul/Fronthaul RTT	$\sim$ 0,450 ms
Backhaul/Fronthaul capacity	1Gbps Ethernet
Ethernet MTU size	1500 bytes
Video Client	VLC v. 2.1.0 with MPEG-DASH
Video File	1080p AVC1 transcoded in 1sec samples

The MEC services are loaded on a node using the LXC framework for providing containerized MEC services. Differentiation of services is mapped to different IP addresses; hence, a video service is running on a container using a different IP address than a simple traffic generator application. Both of these addresses are within the same address space that the UE is using to communicate with the Core Network.

We employ different services in order to measure the performance of the under-study scheme. For video testing, we employ an MPEG-DASH server that streams transcoded videos of up to 1080p resolution, for video segments of 1 sec. This means that the client running on the UE side requests a video segment for the next second that will be played from a selection of available transcodings. Each DASH client requests a Media Presentation Description (MPD) file from the server. According to the descriptions of the available segments and the video requesting algorithm running on the application, the respective segment is requested is downloaded to the client. We use VLC as the end-user application, based on the policies that are described in [16]. The policy that we use is the following: for each video segment, VLC estimates the channel's download rate. For the next segment to be downloaded, it will request the video with coding rate equal to the download rate, if the local buffer status is above 30%. If not, the lowest representation is requested. In the case that the requested transcoding does not exist (since the video coding rate might be significantly lower than the actual channel rate), it will request the next lower representation available. Using this policy we measure the convergence time and estimated channel rates for downloading the best video quality available.

The topology for our experiments is given in Figure 3. The current version of F1oIP is only allowing the data plane split between the CU and the LTE DU. Therefore, the production of two different binary files is not possible. We emulate this type of disaggregated behavior by injecting delay between the network interfaces that are used for this communication be-

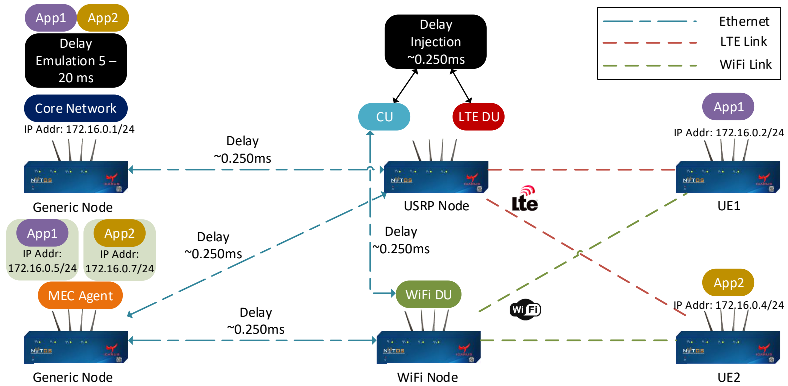


Fig. 3: Experimental topology for the evaluation of the MEC scheme in the NITOS testbed

TABLE II: RTT Results (in milliseconds) for LTE and WiFi access to the service

	LTE to FH	WiFi to FH	LTE to EPC	WiFi to EPC	LTE to EPC (5ms)	WiFi to EPC (5ms)	LTE to EPC (10ms)	WiFi to EPC (10ms)	LTE to EPC (20ms)	WiFi to EPC (20ms)
Avg. RTT	19.7	4.78	32.32	5.26	36.66	9.09	41.58	15.19	51.8	25.14
Min. RTT	15.1	4.39	26	4.59	29	8.6	32.9	14.5	40.8	24.4
Max. RTT	24.7	5.12	43.4	6.64	48.9	9.73	61.9	17	59.9	25.7

tween the CU and DU, equal to ∼0,250ms. The delay injection is done with the *netem* application and is approx. equal to the mean delay that we measure over the fronthaul interface between two nodes of the testbed. Table I is summarizing all of our experimentation parameters.

V. System Evaluation

For the evaluation part of the platform, we focus on measuring two different network performance indicators: 1) the overall latency for accessing the MEC services and 2) the time to converge for streaming high quality video from the service. For both of the setups, we use two multihomed UEs connected to two DUs (one LTE and one WiFi) and measure on the path between the UE and the service.

A. Latency measurement

We compare the latency time for both access technologies between the UE and the service using two different deployments for the service: one being on the fronthaul, with approx 0.250ms delay between the DU and the MEC agent, and one being on the core network. As in typical deployments the core network is not located so close to the edge, we measure the link for the cases of no latency and for tuning the latency for accessing the service. Thus we get an emulated behaviour that the services are deployed at distant servers for typical values of latency (e.g. San Francisco to New York is approx 20ms).

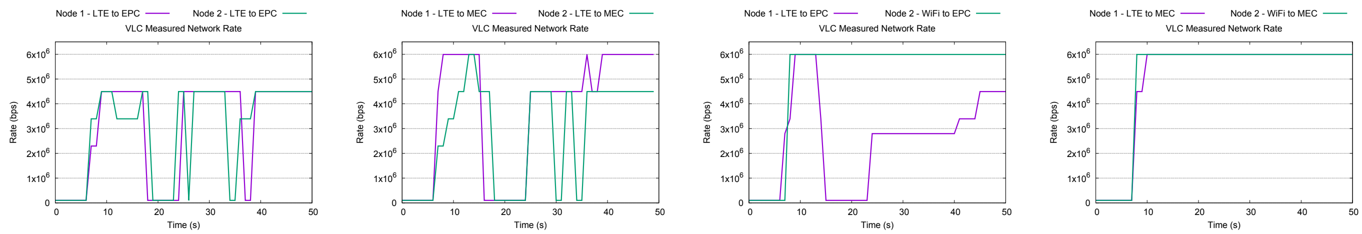
Table II shows indicative RTT times for accessing a service located as a container on the MEC agent or the Core Network (EPC) when accessing the network through either the LTE DU or the WiFi DU. Assuming that latency is almost half of the RTT time, we see that for the cases of MEC access over LTE or WiFi, the latency is consistently less than 10ms, thus allowing several 5G applications to run according to [8]. As we do the experiments in an entirely free from external interference environment, we see that WiFi outperforms the LTE for the cases of latency, even when tuning the delay on the link between the CU and the EPC.

B. Video measurement

For the second part of the evaluation, we test the network with two UEs, connected through either LTE or WiFi and request the video from a server located at the EPC or the MEC server. We plot the requested video rate of the application based on its assumption of the underlying wireless channel, and the current buffer status for the video depicted at the end user. We remind here that for the cases that the buffer status is less than 30%, the minimum representation possible is requested. The plotted video rate is also representing the application‘s perspective on the wireless channel capacity.

Figure 4 shows the results on the selected video rate, and Figure 5 the results on the buffer status of the UE. We see that for the cases that both users use the LTE connection, the selected rates for the application are get barely over 4.5 Mbps. Also, as both users share the same channel, they struggle to get the best video segments that are available and hence their buffer status is kept below 50% for most of the experiment time. When requesting the video from the MEC server over LTE (Figure 4b), one of the two UEs manages to get video rate coded at 6Mbps, whereas the second is bounded at 4.5Mbps, as for the EPC case.

When we use different technologies (one user to LTE, one to WiFi) to request data from the EPC server (Figure 4c), both clients get video coded at 6Mbps, until the LTE UE's buffer is emptied. Then it gradually starts getting better video segments up to 4.5 Mbps. From the other side, the WiFi client quickly converges to getting the best video quality available. For the case of using the same setup to get video from the MEC service, we see that both clients quickly converge to receiving the best available video quality (Figure 4d), and their buffer status is kept full for most of the experiment time (Figure 5d). From these results we conclude that the technology used to request the video plays a key role in the overall experience of the user, whereas the services that are placed on the MEC agent and therefore are closer to the UE outperform the cases of remote testbed placement.



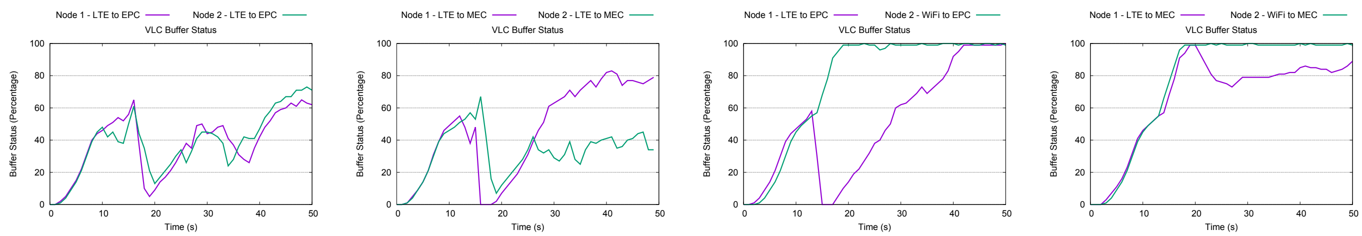
(a) VLC Rate - 2 LTE UEs requesting at the EPC

(b) VLC Rate - 2 LTE UEs requesting at the MEC

(c) VLC Rate - 1 LTE and 1 WiFi UEs requesting at the EPC

(d) VLC Rate - 1 LTE and 1 WiFi UEs requesting at the MEC

Fig. 4: VLC rates for different access technologies



(a) Buffer Status - 2 LTE UEs requesting at the EPC

(b) Buffer Status - 2 LTE UEs requesting at the MEC

(c) Buffer Status - 1 LTE and 1 WiFi UEs requesting at the EPC

(d) Buffer Status - 1 LTE and 1 WiFi UEs requesting at the MEC

Fig. 5: Buffer Status for different access technologies

VI. Conclusion

In this work, we presented a scheme for placing services over the fronthaul interface of heterogeneous 5G base stations. Placing the service on the fronthaul has several benefits as it provides ground for further reduction of the UE to service latency time, and thus support 5G applications with legacy protocols such as LTE. Through a differentiation scheme per each multihomed UE, we can select the technology through which each network client will be served towards even further decreasing the service latency. Our experiments denote that through this technology selection process, and the appropriate placement of services on the MEC, more UEs can be served concurrently with a better Quality of Experience. In fact, the results illustrate better video qualities delivered to the end user of up to 80% by just selecting the radio access technology. In the future, we foresee extending our scheme and adding a machine learning approach on deciding dynamically which services shall be migrated to the MEC server.

Acknowledgment

The research leading to these results has received funding by GSRT, under the act of "HELIX-National Infrastructures for Research", MIS No 5002781 and through the European Horizon 2020 Programme for research, technological development and demonstration under grant agreement N 762057 (5G-PICTURE).

References

[1] "ETSI Multi-access Edge Computing (MEC)," [Online] https://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing.

[2] N. Makris, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas, "Experimental Evaluation of Functional Splits for 5G Cloud-RANs," in *IEEE International Conference on Communications (ICC)*, 2017.

[3] 3GPP, "3GPP TR 38.806 V15.0.0 (2017-12), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study of separation of NR Control Plane (CP) and User Plane (UP) for split option 2; (Release 15)," 2017.

[4] F. Giust *et al.*, "ETSI White Paper No. 24: MEC Deployments in 4G and Evolution Towards 5G," 2018.

[5] N. Makris, V. Passas, T. Korakis, and L. Tassiulas, "Employing MEC in the Cloud-RAN: An Experimental Analysis," in *Proceedings of the 2018 on Technologies for the Wireless Edge Workshop*. ACM, 2018, pp. 15–19.

[6] N. Nikaein, M. K. Marina, S. Manickam, A. Dawson, R. Knopp, and C. Bonnet, "OpenAirInterface: A flexible platform for 5G research," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 5, pp. 33–38, 2014.

[7] ETSI, "ETSI GS MEC 011 V1.1.1 (2017-07): Mobile Edge Computing(MEC); Mobile Edge Platform Application Enablement," 2017.

[8] S. Kekki *et al.*, "ETSI White Paper No. 28: MEC in 5G networks," 2018.

[9] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks - a technology overview," *IEEE Communications surveys & tutorials*, vol. 17, no. 1, 2015.

[10] 3GPP, "3GPP TS 38.473 V15.1.1 (2018-04), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NG-RAN; F1 application protocol (F1AP) (Release 15)," 2017.

[11] ——, "3GPP TS 36.360 V14.0.0 (2017-03), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE-WLAN Aggregation Adaptation Protocol (LWAAP) specification (Release 14)," 2017.

[12] N. Makris, C. Zarafetas, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas, "Cloud-based Convergence of Heterogeneous RANs in 5G Disaggregated Architectures," in *IEEE International Conference on Communications (ICC)*, 2018.

[13] A. Huang, N. Nikaein, T. Stenbock, A. Ksentini, and C. Bonnet, "Low latency MEC framework for SDN-based LTE/LTE-A networks," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[14] C.-Y. Li, H.-Y. Liu, P.-H. Huang, H.-T. Chien, G.-H. Tu, P.-Y. Hong, and Y.-D. Lin, "Mobile Edge Computing Platform Deployment in 4G LTE Networks: A Middlebox Approach," in *{USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 18)*, 2018.

[15] N. Makris, C. Zarafetas, S. Kechagias, T. Korakis, I. Seskar, and L. Tassiulas, "Enabling Open Access to LTE network Components; the NITOS testbed paradigm," in *Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft)*, April 2015, pp. 1–6.

[16] F. Fund, C. Wang, Y. Liu, T. Korakis, M. Zink, and S. S. Panwar, "Performance of dash and webrtc video services for mobile users," in *2013 20th International Packet Video Workshop*, Dec 2013, pp. 1–8.