Employing MEC in the Cloud-RAN: An Experimental Analysis

Nikos Makris, Virgilios Passas, Thanasis Korakis
University of Thessaly
Volos, Greece
{nimakris,vipassas,korakis}@uth.gr

Leandros Tassiulas
Yale University
New Haven, USA
leandros.tassiulas@yale.edu

ABSTRACT

5G network access is expected to deliver high performance with low-latency network connections for the end-users, suitable for a plethora of different applications, as well as add up to the network flexibility and manageability from the operator's perspective. In order to achieve low-latency, Multiple-access Edge Computing (MEC) is considered, whereas for achieving flexibility, the disaggregation of the base station elements and moving parts of their functionality to the Cloud is proposed. In this paper, we consider the case of disaggregated base stations based on the CU-DU paradigm, able to provision MEC functions in a per-packet and per-client basis, over real networks. We evaluate the placement of the MEC functions over the fronthaul interface or collocating them with the Core Network. We employ the OpenAirInterface platform and evaluate our MEC solution with dynamically adaptive video streams. Our results show significant gains for the service-to-UE path latency, complying with the requirements set for the 5G MEC operation.

ACM Reference Format:
Nikos Makris, Virgilios Passas, Thanasis Korakis and Leandros Tassiulas. 2018. Employing MEC in the Cloud-RAN: An Experimental Analysis . *In 2018 Technologies for the Wireless Edge Workshop (EdgeTech'18:), November 2, 2018, New Delhi, India. ACM, New York, NY, USA*, 5 pages. https://doi.org/10.1145/3266276.3266281

1 INTRODUCTION

The ground-breaking network performance about to be delivered by the 5G networks has created fertile ground for multiple novel services and applications. Several different services are used to evaluate the proposed solutions, with the most notable being Ultra Reliable and Low Latency Communications (URLLC), massive Machine Type Communication (mMTC) and enhanced Mobile Broadband (eMBB) [1]. These applications cannot be supported by a single technological solution, but by a suite of protocols that collaboratively delivers the desired end-user experience. To this aim, the New Radio (NR) interface has been introduced [2], coupling several novel technological features, such as functional splitting of the base station stack, new signaling, etc. These features can lower the latency over the wireless channel, by using significantly higher channel bandwidth and accessing new wireless spectrum, such as the cm- and mm-Wave bands. The development of 5G-NR is also taking into account the emerging trend for network virtualization, through enablers such as NFV-MANO [3] for enhanced network management, ready for Multi-access Edge Computing (MEC) [4] that can efficiently serve low-latency applications. Multi-access Edge Computing is the transformation of the legacy Mobile Edge Computing that includes other access technologies as well.

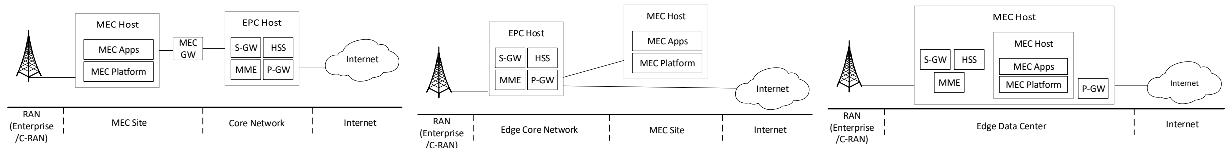
In this paper, we consider such a 5G environment, and argue from the architectural perspective on the solutions for delivering a fully-fledged MEC solution for disaggregated multi-RAT base stations. We reckon the functional disaggregation of the base station unit at the high Layer 2 of the OSI stack (between the PDCP and RLC layers), as denoted in the 5G-NR standards [5]. The entities incorporating the RLC, MAC and PHY functionality are hereafter mentioned as Distributed Units (DUs), whereas the higher layers with PDCP, RRC and Core Network interfaces are mentioned as Centralized Units (CUs). The latter ones can be instantiated in the Cloud. In such a disaggregated setup, with no stringent requirements for latency between the CUs and DUs for the fronthaul, the transport network can be realized by using packet-based technologies (e.g. Ethernet, e-CPRI), whereas the DUs may also implement different technologies for the wireless network access e.g. 5G-NR, LTE or WiFi.

This work is providing the architectural substrate, signaling and experimental analysis of a MEC framework running on the fronthaul of a C-RAN setup. The main contributions of this paper are:

- to provide a fully-fledged solution for directly plugging new services on the fronthaul network of disaggregated network setups,
- to provide the needed signaling functions for orchestrating fronthaul edge services,
- to experimentally evaluate the framework and provide numerical evidence on its efficiency and the end-user performance for dynamic adaptive video streaming services.

EdgeTech'18: , November 2, 2018, New Delhi, India

ACM ISBN 978-1-4503-5931-3/18/11…$15.00
https://doi.org/10.1145/3266276.3266281



(a) Bump in the wire; MEC service is placed over the S1 interface

(b) MEC placement after the Edge located Core Network

(c) MEC is placed between the distributed EPC entities

Figure 1: Different types of MEC deployments for 4G and beyond networks

2 MOTIVATION AND RELATED WORK

The disaggregation of base stations to multiple entities is a key technology for 5G networks. Several options for splitting the base station units have been proposed [6] at different layers of functionality. Ideally, the wide capacity that optical fiber links offer can be used for rigorously transferring data between the disaggregated elements, allowing the low layer splitting of the stack, even at the baseband. Nevertheless, this requires low latency, jitter and constant high throughput fronthaul connections. As an answer to this, splits have been identified for the higher layers of the stack as well [7], with lower demands on the fronthaul latency, exploiting Ethernet based connections. In [8], the authors experiment with such Ethernet based encapsulation for the fronthaul, for two different types of splits; MAC/PHY and PDCP/RLC. In the [9], the work is extended to include new signaling for the data plane communication between the PDCP and RLC layers of the stack. This signaling is referred as **F1 over IP (F1oIP)**, as it resembles the standardized F1AP [10]. Since these splits use Ethernet encapsulation, they facilitate the introduction of services in the fronthaul.

C-RAN deployments add to the network flexibility, with the dynamic switching of technologies for serving the UEs, instantiation of new base stations in an area based on the demand, etc. On the other side, MEC can significantly reduce the service-to-UE latency, allowing time critical services to be offered over the cellular network. ETSI white-paper on MEC [11] is providing an overview of the different types of deployment in 4G and beyond networks. Fig. 1 is providing these methods: in the "bump in the wire" method (Fig. 1a) the MEC service is being placed between the RAN and the Core Network, by means of a proxy that overhears the S1-U packets and handles them in the case they are destined for it. In the collocation with the Core Network case (Fig. 1b and 1c), MEC is placed just after the PGW interface of the Core Network, or among the disaggregated Core Network elements. These two solutions require the Core Network to be deployed at a data center close to the network edge.

The benefits of MEC have been widely analyzed in relevant literature. In [12], a survey on the evolution of the MEC solutions is presented, and some insights are given on the integration of MEC with NFV-MANO. In [13], authors present their solution for a *Turbo charged edge*, able to boost the delivered throughput for dynamic adaptive video streaming applications about 30%. In this work, we employ a Cloud-RAN architecture based on the CU/DU split options, with the goal to place the MEC functions over the fronthaul. A similar study that resembles this approach is [14], where the benefits of a co-deployment between the CU and a MEC service are analyzed in terms of networking, exposed information and security. Although the study is considering the CU/DU split for 5G networks, the MEC function is placed just after the CU side of communication, thus realizing the bump in the wire approach (Fig. 1a) closer to the edge.

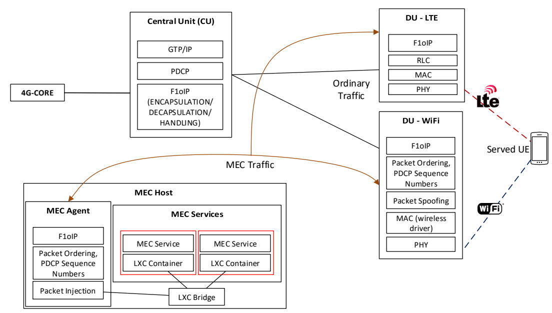


Figure 2: MEC placement on the fronthaul interface of the CU/DU split

3 MEC ARCHITECTURE

In this section we detail the different components needed for realizing the MEC solution over the fronthaul interface. We use as our starting point the F1oIP implementation of the CU/DU split, detailed in [9], and extend it accordingly in order to enable the MEC operation. The implementation is introducing a new signaling mode between the PDCP and RLC layers, resembling the F1AP standardized interface for the communication between CUs and DUs. Originally, F1AP is handling data plane packets over GTP tunnels, established for each served UE of the system. The F1oIP implementation is using UDP/TCP interfaces in order to exchange the traffic between the CUs and DUs, including also some signaling information on the packet headers that is ordinarily exchanged between the PDCP and RLC layers (e.g. DRB/SRB allocation, frame/subframe scheduling, protocol context, etc.). The following sections describe the new packet headers that are introduced over the F1oIP interface, and how they allow applications directly connected to the fronthaul interface handle data-plane user-directed traffic.

3.1 Disaggregated base station communication

According to 5G-NR, the disaggregated functionality of the 5G base stations addresses several technologies. To this aim, the proposed standardized split option by 3GPP resides in the high layer 2 of the OSI stack, between PDCP and RLC. As this split option has more slack limitations on the latency and throughput over the fronthaul (less than 100Mbps for serving a single 20MHz base station unit), different technologies can be used for the DU side of communication. Target technologies to be supported by DUs are 5G-NR, LTE, WiFi and their evolution. The relationship between CUs and DUs is 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.

We use as a starting point the implementation of F1oIP, which handles the communication between the CU and heterogeneous DUs of the system. In this implementation, each packet exchanged over the fronthaul interface is bearing on its header scheduling information to be used by the lower layers, according to Fig. 3. Based on this information, the packet is assigned to the respective transport channels of RLC and is then left to MAC layer for scheduling its transmission over the air. For the case of non-cellular DUs, the respective information is not handled from the respective DU software. For example, in the case of a WiFi DU, the F1oIP header information related to the scheduling of the packet is ignored. For the UL case, the reverse process takes place before transmitting the packet to the CU. This means that the DU is assigning new PDCP headers and adds the respective information on the header in order for the packet to be handled at the CU side.

3.2 DU-MEC communication

In order to place the MEC functions over the fronthaul network, we employ a similar protocol as for the CU-DU communication, by means of a **MEC agent**. This software is in charge of generating and transmitting the relevant messages destined to the DU for the DL case, and receiving and delivering traffic destined to the MEC service from the DU for the UL case. We introduce two functions for the MEC offloaded traffic: whenever the DU side of the system receives traffic destined to the MEC platform, it spawns a *mec_data_request* message. This message is destined to an agent service providing the bridge to the MEC functions. More details on the differentiation of the MEC services is following in the next subsection. Since at this layer each traffic flow is differentiated based on the RNTIs assigned to the UEs, this information can be exploited for selecting which clients shall use the service. For the reverse path (MEC service sending traffic to the UE), the MEC agent sends a *mec_data_indication* message to the DU serving the client. This process is agnostic of the technologies that serve the end user (5G-NR, LTE, WiFi), as long as the DU can handle these messages.

Since the data plane path of the base station is split between the PDCP and RLC layers, and MEC functions are introduced in-between the path, the DU side needs to be aware of the similar scheduling information that is sent from the PDCP layer. To this aim, we introduce a new message exchange sequence between the DU and the MEC agent, for updating the scheduling information on the agent (e.g. DRB used, client mapping to RNTIs, etc.). Moreover, the agent is in charge of generating and assigning PDCP sequence numbers and encapsulating the MEC data in PDCP frames.

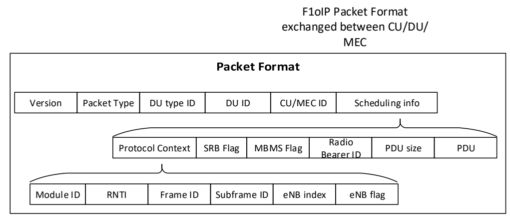


Figure 3: F1oIP packet format exchanged over the fronthaul between the CU, DU and MEC agent

3.3 Support for multiple MEC services

The crucial part of the platform that allows the offloaded MEC functions to take place is the MEC agent. This software is orchestrating the communication of the MEC services with the DU side of the system, and delivering the traffic to the services running on top of it. Whenever the agent receives MEC intended traffic, it decapsulates and injects it to the MEC service. We select to run the MEC services as Linux Containers (LXC), as they can be instantiated on the fly, whenever an end-user requests different services from the MEC platform. Employing LXC containers has multiple benefits; it allows each new service to be addressed with a new container, with a new network IP address, that 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 service containers 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 implementing the MEC service.

4 EXPERIMENTAL SETUP

In this section we detail our experimental setup and methodology. The described functionality has been developed over the OpenAirInterface platform (OAI) [15], that provides an open source implementation of the base station stack that can be executed over commodity hardware. We conduct the experiments over the NITOS testbed [16] that offers a rich remotely accessible experimentation environment with resources spanning from LTE to WiFi and Software Defined Radios that suit our experimentation needs.

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 module developed in [9] in order to setup a separate communication channel between each DU and the MEC Agent. This channel, and the F1oIP channels for the CU/DU communications are selected to be TCP over Ethernet, as our former experiments denote that there is no notable performance degradation compared to UDP or even the vanilla OAI setup. For the MEC configuration inside the DU, the new messages are sent after checking the client's information.

Each client at this layer of handling is represented with an RNTI value. At the DU side, we have a mapping of the IP addresses of each client to RNTIs in order to differentiate the clients that are using the MEC service with the ones who do not. This means that the DU can be aware of which client's traffic shall be offloaded to the MEC service, without performing any Deep Packet Inspection techniques.

Table 1: Equipment parameters

Network Parameters	Values
LTE mode	FDD Band 7
LTE Frequency	2680 MHz (DL)
Antenna Mode	SISO
No RBs	50 (10 MHz)
UE	Cat. 4 LTE, Huawei E3272
Backhaul/Fronthaul RTT	~ 0,250 ms
Backhaul/Midhaul capacity	1Gbps Ethernet
Ethernet MTU size	1400 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. For every packet that the MEC Agent is receiving, it is deserialized from the F1oIP protocol and injected to the containers providing the services. Different services can be addressed to different containers. The under study service is an MPEG-DASH server, able to stream transcoded videos of up to 1080p resolution, for video segments of 1 sec. The server is running over an Apache2 web service, in the MEC containers and the Core Network for comparing their performance. Each DASH client requests a Media Presentation Description (MPD) file from the server. According to the descriptions of the available video segments and the video requesting algorithm running on the application, the video is downloaded to the client. We use VLC as the end-user application, based on the following policy: for each video segment, VLC estimates the channel's download rate and for the next segment it requests the video sample with coding rate equal to the download rate. In the case that it does not exist (since the video coding rate might be significantly lower than the actual channel rate), it requests 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 Fig. 4. The current F1oIP version 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 disaggregated behavior by injecting delay in the network used for the CU - DU communication, equal to ~0,250ms, which is the mean delay that we measure over the fronthaul. We omit the incorporation of the WiFi DU in the system, as we want to focus on the LTE network operating in licensed spectrum, as the environment is entirely interference free. Table 1 is showing our experimentation settings.

5 SYSTEM EVALUATION

In this section we present our experimental findings. We evaluate our MEC over Fronthaul scheme by placing the MEC

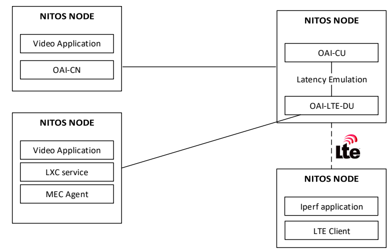


Figure 4: Experiment mapping over the NITOS testbed

Agent over the fronthaul interface and compare it against the solution of deploying the service on an edge-located datacenter with the service placed after the PGW component (see. Fig. 1b). Prior to measuring the video delivery performance, we assess the jitter and latency for the two different schemes. We use the *iperf* measurement tool for generating UDP traffic equal to the maximum data volume exchanged when streaming a 1080p video file (8Mbps).

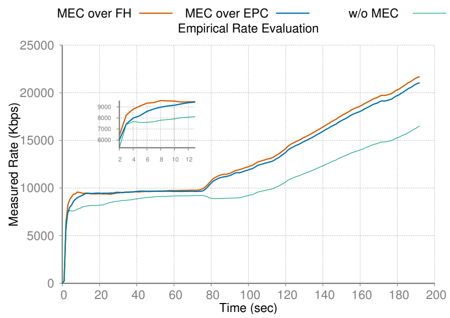
Table 2: Latency and Jitter Results

	MEC over FH	MEC over EPC
Avg RTT	19.919 ms	36.254 ms
Min RTT	15.785 ms	31.825 ms
Max RTT	22.713 ms	42.553 ms
Avg Jitter	0.414 ms	0.527 ms
Min Jitter	0.409 ms	0.514 ms
Max Jitter	0.421 ms	0.541 ms

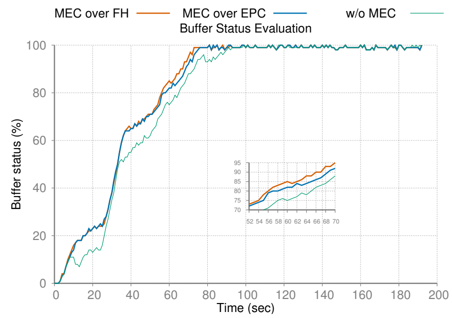
Table 2 shows the reported values averaged after 10 experiment runs for the path UE - RAN - Service. We see that in all the cases, Round Trip Time (RTT) and jitter are lower for the MEC over Fronthaul case compared to the EPC one. The results are very encouraging given the 5G network requirements for latency of time critical applications. Since latency is about half of the RTT, placing the MEC functions over the fronthaul equals to network latency less than 10 ms, which is the target for several applications (e.g. V2X, Smart Factory, e-Health, entertainment) as denoted in [14].

As a second set of experiments, we investigate the behaviour of video streaming, using application layer metrics, for the two cases of placing the MEC function against not using MEC and setting up an end-to-end path with latency in the ranges of approx. 30 ms. We plot: 1) the VLC reported empirical rate, meaning the perception of the application of the end-to-end network (server to application), and 2) the buffer occupancy status for displaying the video to the end-user. These two metrics are essential for the end-user Quality of Experience (QoE); due to the policy used for requesting the next video segment, the requested video rate equals to the empirical rate (or less if there is no such representation).

Figure 5 is plotting our measured metrics. The results are averaged of 10 experiment runs for each case. We see that the pattern for the empirical rate is exactly the same for the two MEC functions, for requesting the same video of duration approx. 190 secs. In the case of placing the MEC service over the fronthaul interface, the application understands the network channel as being of better quality, constantly



(a) VLC reported empirical rate



(b) VLC reported buffer status

Figure 5: Experimental findings from running MPEG-DASH

reporting approximately 1Mbps higher than the MEC over EPC case (see Fig 5a). The convergence time for requesting the best video quality available is also notable (during the first 20 secs of the video). As we see, in the first 10 secs of the video (equal to the first 10 requests for video samples), the MEC over fronhaul case is reporting more than 2Mbps of perceived channel quality (zoomed plot in Fig. 5a). This equals to a much better quality of the video presented to the end user in this time period, as the MEC over fronthaul almost immediately requests the higher representation available (1080p). Comparing these results with the no MEC case, we see that the perceived channel quality is even worse, by approx. 3Mbps for the entire duration of the experiment. For the case of buffer status, we observe that both MEC representations follow the same pattern. For the fronthaul MEC case, the buffer is more filled by at least 5%, although the quality of the video is better, meaning that MEC over fronthaul offers video of better quality and for longer buffered periods.

6 CONCLUSION

In this paper, we presented our solution for placing MEC over the fronthaul interface of a Cloud-RAN setup formed according to the standards for 5G-NR. We introduced new signaling for the communication between the system's DUs and a MEC agent that provides access to services running over containers in the fronthaul. Our proof-of-concept results showed that the latency achieved is complying with the requirements of several 5G applications, designed to run on the network edge. We evaluated our solution for MPEG-DASH adaptive streaming for the cases of placing the MEC functions over the fronthaul or collocating it with the Core Network. Our results show that our solution in the former case is able to serve the end-user with better video quality for longer buffered segments. In the future, we foresee to extend this framework for user mobility. As V2X is a use case for 5G applications, we will examine how the service replication over different edge nodes can serve UEs which rapidly handover between base stations of different technologies.

ACKNOWLEDGMENT

The research leading to these results has received funding by GSRT, under the act of "HELIX-National Infrastructures for Research", MIS no 5002781.

REFERENCES

[1] 3GPP TR 22.864 V15.0.0 (2016-09), 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers - Network Operation; Stage 1 (Release 15), author = 3GPP, year = 2016,.

[2] 3GPP. "3GPP TS 38.401 V0.2.0 (2017-07), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NG-RAN; Architecture description (Release 15)", 2017.

[3] ETSI. "ETSI GS NFV-MAN 001 V1.1.1 (2014-12), Network Functions Virtualisation (NFV); Management and Orchestration ", 2014.

[4] "Multi-access Edge Computing". [Online],https://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing.

[5] 3GPP. "3GPP TS 38.470 V0.3.0 (2017-09), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NG-RAN; F1 general aspects and principles (Release 15)", 2017.

[6] 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*, 17(1), 2015.

[7] I. Chih-Lin, J. Huang, Y. Yuan, S. Ma, and R. Duan. "NGFI, the xHaul". In *Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2015.

[8] N. Makris, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas. "Experimental Evaluation of Functional Splits for 5G Cloud-RANs". In *IEEE ICC*, 2017.

[9] 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 ICC*, 2018.

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

[12] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella. "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration". *IEEE Communications Surveys & Tutorials*, 19(3):1657–1681, 2017.

[13] I. Chih-Lin, S. Han, Z. Xu, S. Wang, Q. Sun, and Y. Chen. "New paradigm of 5G wireless internet". *IEEE Journal on Selected Areas in Communications*, 34(3):474–482, 2016.

[14] A. Reznik et. al. "ETSI White Paper No. 23: Cloud RAN and MEC: A Perfect Pairing", 2018.

[15] 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*, 44(5):33–38, 2014.

[16] "NITOS - Network Implementation Testbed using Open Source platforms.". [Online],https://nitlab.inf.uth.gr/NITlab/.