Multi-Access Edge Computing Architecture Optimized for Performance Driven Radio Access Networks in 6G Era

Gee-Kung Chang, Yahya Alfadhli, You-Wei Chen
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
geekung.chang@ece.gatech.edu

Abstract: Perspective and challenge of the MEC implementation, merging with the RAN architecture for 6G mobile networks are discussed from futuristic use-cases point-of-view, including mobile operators and application developers. Featuring demonstrations with AI/ML are also highlighted.

OCIS codes: (060.4256) Networks, network optimization; (060.4258) Networks, network topology.

1. Introduction

Multi-access edge computing (MEC) can be defined as the placement of some of cloud-computing capabilities at the edge of the network [1]. MEC was initially introduced as a fog computing concept to cope with the increased number of internet of things (IoT) devices that should be processed locally to alleviate the load at the core of the network. The prime benefit of MEC is that it reduces the end-to-end latency for latency sensitive applications which has become an essential enabling feature for 6G networks to enable different futuristic use-cases. However, the distribution and virtualization of RAN functions as well as the rapid advances of artificial intelligence and machine learning algorithms have led to a unique paradigm wherein MEC will penetrate every layer in 6G systems. The distributed RAN architecture is represented in three geographical units and two links: central unit (CU), distributed unit (DU), remote unit (RU), fronthaul link (FH) and midhual link (MH). Figure 1(a) depicts the scheme of the envisioned MEC-based RAN architecture wherein, MEC is implemented across the continuum of RU, DU, and CU.

In this paper, we focus on the MEC design from the perspective of the major users: application developers and mobile operators. While mobile operators can give access to trusted third-party application developers to design innovative applications and services for end users, they themselves can benefit from the available computational resources to improve the quality of user experience and to increase the efficiency of their system. Therefore, we discuss some potential use-cases for both application developers and mobile operators.

2. MEC use-cases for application developers

There are three important factors that impact the design and resource allocation for the distributed MEC entities, which are: the scope of data context, latency tolerance of the application, and the centralization of the MEC resources. First, the scope of the data context can be defined as the scope at which data is being generated and consumed. There are four data context scopes including intra-RU, intra-DU, intra-CU and break-out to cloud, as depicted in Fig. 1(a).

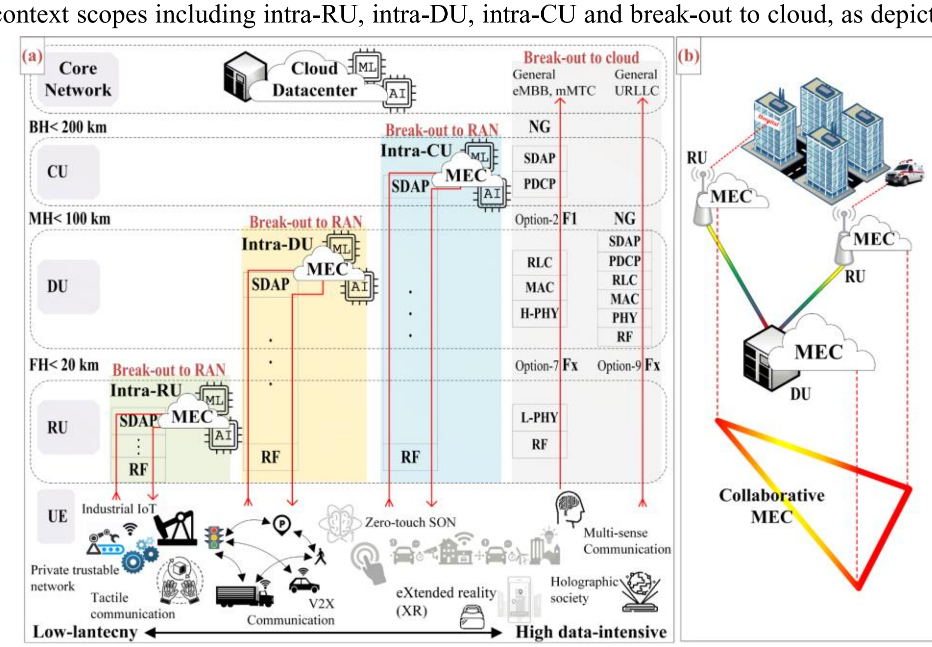

Fig. 1: (a) 6G RAN architectures for different application context scopes including cloud and RAN breakouts (b) collaborative MEC example

For example, in intra-RU context scope, data is being generated and consumed within a single RU as opposed to intra-DU, where data is being generated at one or multiple RUs and consumed at the same or different RUs but within the same DU. Consider the augmented reality (AR) assisted driving scenario, where data is generated by multiple street cameras and cars sensors, processed by the MEC unit and sent back to the car. In this scenario, all the input data for the MEC is available within a single DU, which makes the application design much simpler. Moreover, the input video streams from cameras, if not needed by some other applications, are irrelevant for users outside of that DU scope. Hence, to avoid wasting of network resources, data should be kept within their application context scope by having a RAN breakout. This is similar to the local breakout concept, which is one of the major functions of MEC that enables operators to steer the traffic in their network [1]. The most right vertical box in Fig. 1(a) corresponds to the scenario where RAN-breakout is not used and data traverses the whole RAN network to the cloud. This general case is suitable for less demanding ultra-reliable low latency (URLLC), enhanced mobile broadband (eMBB), and massive machine type communication (mMTC) applications.

Second, is the latency tolerance of the use-cases or applications, which is tightly related to the physical lengths of FH/MH/backhaul (BH), and the network topology and load conditions. Applications with limited latency tolerance constrain the flexibility of MEC placement in the network. Hence, for mission critical and real-time applications that aim to achieve 1 ms end-to-end latency, MEC entity should be placed at the RU especially in cases where the fronthaul is too long or has high latency/jitter, or in cases where DU and RU are placed in the same geographical location.

Third, is the centralization of MEC resources, which is important for two main reasons. (*i*) It reduces the cost and complexity of the system. MEC is built on costly hardware (*i.e.*, servers, storage, CPUs/GPUs, etc.) which increases the CAPEX/OPEX of the system. Hence, it is preferred to centralize the MEC hardware as much as possible to increase the multiplexing sharing gain. (*ii*) From the application developer prospective, centralized MEC corresponds to higher computational power and more available input data. On the other hand, an alternative method to centralization is to use collaborative MEC, as depicted in Fig. 1(b). In this scenario, MEC units can barrow some resources from other neighboring MEC units residing in other RAN units, which leads to higher utilization and sharing gain.

In order to allow the MEC centralization and at the same time satisfy latency sensitive applications by maintaining the 1 ms end-to-end latency goal, enhanced common public radio interface (eCPRI) defines the fronthaul requirement for low latency applications as "Class high25", where the one-way latency is limited to 25 us [2]. This latency budget is merely sufficient for 5 km optical fiber propagation delay. However, all digital based function split options inherently introduce latency at the fronthaul, limiting the fronthaul length. Hence, a possible solution to reduce fronthaul latency is to use analog radio-over-fiber (A-RoF) based fronthaul, also known as Option-9, wherein all PHY processing, including RF layer, is consolidated at the DU, and RU only performs optical/electrical conversion [3]. A-RoF based fronthauls had been challenged by the non-linearity impairment which is directly proportional to distances above 10 km [4]. However, this is no longer an issue as the future-proof fronthaul length is latency-limited to 5 km.

3. MEC use-cases for mobile operators

Figure 2(a) illustrates one example of MEC host implementation. Basically, MEC virtually resides on physical hardware to provide several useful services, including, but not limited to, localization, radio network information system (RNIS), artifactual intelligence (AI), machine learning (ML), flexible radio access network (RAN) manager, big data and self-organized network (SON). All these services can be accessed by application developers and mobile operators through common application programming interface languages (APIs).

The distributed nature of 5G RAN provides a suitable habitat for MEC hardware. However, the usefulness of MEC and its various services, makes MEC stands out as the heart of 6G networks. These two concepts can be seen clearly from Fig. 2(b). At the heart of these three RAN units, we can see the MEC units interfacing with all other components at any particular unit. Moreover, MEC entities from different RAN units can be collaboratively communicating to further increase resources utilization and enable collaborative-based applications. Hence, MEC can be interpreted as the platform wherein users such as mobile operators, application developers, and content providers can interface with resources such as RAN functions, FH interface (F1), MH interface (Fx) and BH interface (NG). Hence, we discuss

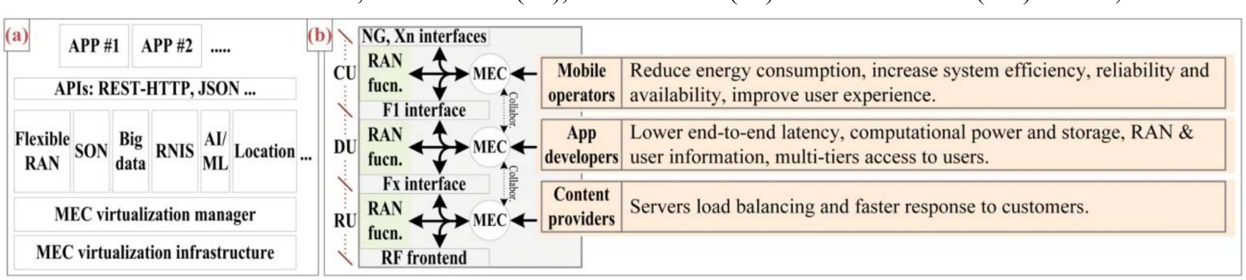

Fig. 2: (a) MEC host example; (b) Multi-tier collaborative MEC paradigm in 6G RAN architecture.

the usage of MEC resources for enhancing network performance and user experience including flexible RAN manager and optimizing the low layers of RAN (*i.e.,* MAC/PHY/RF).

3.1. MEC for flexible RAN management

RAN architecture has been evolving through different mobile network generations. Moving to 5G, the regular base station, known as eNodeB/gNodeB (eNB/gNB) station in 3GPP standard, can be split into different units (*i.e.,* CU, DU, RU) depending on the deployment scenario. Splitting of RAN functions was originally motivated by several factors such as reducing the cost and complexity of the base stations, reducing FH bandwidth and latency requirements and allowing enough centralization degree to enable the integration between 4G, 5G, 6G and other non-3GPP access networks such as WiFi. The topic that studies the splitting and arrangement of RAN functions is known as function split analysis or RAN functional decomposition and is thoroughly surveyed in [5]. The 3GPP standard defines eight main function split options, each of which has distinct merits and requirements [6].

Several resources in literature, including some prior work [7], investigate the concept of flexible function split wherein the network can dynamically change the function split over time. For instance, Option-7 can be used under normal fronthaul traffic conditions. Then, once the network is getting congested, the flexible network will switch to a less bandwidth-demanding function split such as Option-6. Transition between splits should be smooth to avoid service disruption and degradation. Efficient predictive algorithms can facilitate achieving smooth transition in a proactive manner. Optimized placement of functions in the flexible function split network is an interesting topic that has been investigated heavily in literature [8]. Such optimization algorithms can be considered as a service of MEC. Beside flexible RAN management service, several network optimization measures can also be facilitated by MEC, including, but not limited to, controlling energy consumption at RU, balancing computational loads among the network entities, and maximizing network resources utilization.

3.2. MEC for optimizing the low layers of RAN

Self-optimization, self-configuration and self-healing are the three major components of SON paradigm introduced by 3GPP standard. These functions collectively promise mobile operators to improve network performance and user experience while reducing capital and operational expenses. SON and MEC are similar in two main aspects: (*i*) Their functions can be distributed among different RAN units (*i.e.*, CU, DU). (*ii*) They rely on continuous feedback and system monitoring. This makes SON a suitable candidate to be considered as an MEC service, which can realize several standard SON use-cases such as: (*i*) Automatic neighbor relation (ANR): generating configurations for neighboring radio elements to support mobility, load balancing and dual connectivity. ANR can manage relations between different gNBs, eNB and beams, which enables dynamic resource management. (*ii*) Enhanced inter-cell interference cancellation (eICIC): aiming to enhance the performance at the edge of the cell by mechanisms such as coordinated multi point transmission (CoMP) and fractional frequency reuse (FFR). (*iii*) Mobility robustness optimization (MRO): optimizing handoff configurations to improve user experience and save network resources.

Beside these defined SON functions, plethora of innovative algorithms based on machine learning and artificial intelligence are developed to solve some challenging problems such as interference mitigation and reduced system efficiency [9]. Interference, whether artifactually generated by an attacker or naturally caused by nearby cells or users, can reduce system efficiency and cause service degradation. For instance, [10] exploits SARSA reinforcement learning to solve the interference problem. Another work, [11], attempts to solve the multi-user interference in an analog radio-over-fiber (RoF)-based fronthaul using multi-level artificial neural network (ANN) non-linear equalizer. Similarly, authors of [12] discuss the design of deep neural network (DNN) decoder for digital signals recovery with flexible modulation formats. Lastly, authors of [13] use Q-learning for energy-efficient resource allocation in 5G RAN. Their simulation results show benefits such as increasing energy and spectral efficiencies and mitigating the interference.

4. Conclusion

While highly distributed MEC placement is inevitable for certain real-time applications, it imposes economical and operational challenges. MEC centralization can be achieved physically by the use of low latency fronthaul based on A-RoF technology, or logically by allowing multiple MEC units to collaborate. We also discussed the importance of MEC in 6G RAN for improving the performance for both mobile operators and end users.

References

[1] S. Kekki, et al., MEC in 5G-ETSI White Paper No. 28 (2018)
[2] "Common Public Radio Interface: Requirements for the eCPRI Transport Network", IEEE Standards Association, V1.2, (2018).
[3] Y. Alfadhli, et al., Comput. Netw., vol. 162, 106865 (2019).
[4] W. P. Ng, et al., IEEE Photonics J., 4(5), (2012), pp.1560-1571.
[5] L. Larsen, et al., IEEE Commun. Surv. Tutorials., 21(1), (2018).
[6] "Radio access architecture and interfaces." 3GPP TR 38.801, (2016).
[7] Y. M. Alfadhli, et al., in *Proc. OFC* (2018), paper Th2A-48.
[8] A. Marotta et al., in *Proc. ICC* (2018), pp. 1-6.
[9] Y. Sun, et al., IEEE Commun. Surv. Tutorials., 21(4), (2019).
[10] Q. Zhou, et al., Opt. Lett., 44(17), (2019), pp. 4347-4350.
[11] S. Liu, et al., IEEE PTL, 30(19), (2018), pp. 1675-1678.
[12] Q. Zhou, et al., Opt. Lett., 30(17), (2018), pp. 1511-1514.
[13] I. AlQerm, et al., in *Proc. IEEE PIMRC* (2017), pp. 1-7.