The Role of Cloud and MEC in 5G

Fabrizio Granelli

Abstract 5G will incorporate several state-of-the-art architectural and protocol approaches on top of his networking infrastructure to tackle the emerging needs for improved flexibility and performance.

Such requirements lead to the design of two options for the 5G architecture: one, which represents an evolution of 4G LTE standard IP architecture, and the other where the core network functions interact with each other using a Service Based Architecture (SBA). This novel architecture allows the integration of the recent developments in the field of Cloud technology and Mobile Edge Computing. This chapter will introduce how the cloud approach and edge computing paradigm can be integrated in the upcoming 5G standard for next generation cellular networks.

1 Introduction

5G is expected to support unprecedented requirements. Indeed, besides a predictable increase in data transfer performance and spectral efficiency, 3GPP 5G requirements include very low latency (in the order of msec), high reliability, capability to offer access to distributed computation and storage facilities in addition to connectivity and bandwidth.

Knowing the complexity of current 4G/LTE architectures, it seems extremely hard to define a single architecture able to satisfy ALL the expected performance requirements. Nevertheless, several target scenarios and services include strict constraints in terms of low latency and extremely reliability. Those services can be classified as URLLC (Ultra-Reliable Low Latency Communication) services. Vehicular communications and remote control of robots or machinery belong to such class of services, and they are well-known 5G application scenarios.

Fabrizio Granelli
Consorzio Nazionale Interuniversitario per le Telecomunicazioni - Sede di Trento, Via Sommarive 9, Trento, ITALY, e-mail: fabrizio.granelli@unitn.it

As a consequence of the above issues, it is necessary for the 5G architecture to incorporate additional functionalities, by exploiting the recent solutions existing on the Internet: the Cloud and virtualization.

Cloud Computing represents the current paradigm for the delivery of services to a massive number of users, and it is based on the concentration of huge computation and storage resources in strategic locations (the datacenters) that can run services in an efficient and scalable manner by using advanced management and virtualization approaches.

Virtualization represents a relatively old paradigm for enabling access to shared resources, that in the modern days gained enormous attention due to its successful application in support of cloud computing, but more recently also in the area of networking. Indeed, by virtualizing key network functionalities, it is possible to detach software functions from dedicated hardware, with the advantage to be able to re-locate or modify their resources most efficiently and in real-time. This emerging paradigm is defined Network Function Virtualization (NFV).

The above concepts can be introduced in the design of the next generation of mobile networks, leading to the definition of two emerging paradigms in the design of 5G: Cloud Radio Access Network (Cloud RAN) and Mobile Edge Computing or Mobile Edge Cloud (MEC). The purpose of this chapter is to provide some basic information of such paradigms as well as their relationship with the 5G standard.

2 The Cloud Radio Access Network paradigm

A relevant issue in the design of next generation mobile networks is related to the fact that, in order to enable the wireless communication technologies to provide high performance to the mobile users, cells are increasingly becoming smaller and smaller. This generates problems in terms of costs for buying the equipment as well as powering them, since currently (in 4G) cellular Base Stations are extremely expensive.

In this framework, the Cloud RAN (C-RAN), sometimes also referred to as Centralized-RAN, represent a possible architecture for future cellular networks to tackle the issues of costs and energy consumption [1].

Indeed, Base Stations are designed to handle the maximum traffic, not average traffic, resulting in a waste of processing resources and power at idle times or in situations of low amount of traffic. The problem is related to the fact that the majority of power does not scale with the number of users or the amount of traffic, being related to power the RF interfaces and cooling the processing elements. Therefore, a more flexible solution is required.

The typical RAN architecture in 4G/LTE is depicted in Figure 1, where the Base Stations contain both RF and signal processing functionalities in the so called "evolved Node B" (eNodeB) architecture. Cloud RAN introduces a detachment between the RF interfaces (called Remote Radio Heads - RRHs) and the signal and data processing functionalities (called Base Band Units - BBUs), leading to an

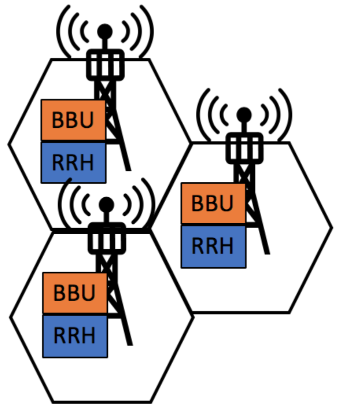

Fig. 1 Typical RAN Architecture: the eNodeB contains both RF and processing capabilities.

architecture like the one presented in Figure 2. To some extent, C-RAN may be viewed as an architectural evolution of a distributed base station system, that takes advantage of many technological advances in wireless, optical and IT communications systems. Moreover, Cloud RAN is a paradigm whose scope is limited to the 5G Radio Access Network (RAN), and especially is impacts mostly the eNodeB architecture.

In Cloud RAN, the detachment of BBUs from the RRHs is achieved by means of the latest CPRI standard, that enables to interconnect the two components using optical fibers or wireless through the fronthaul links. Optical trnasmission is achiwved through Coarse or Dense Wavelength Division Multiplexing (CWDM/ DWDM) technology, while mmWave communications are used to allow transmission of baseband signal over long distance.

BBUs are then grouped together by applying recent Data Centre Network technology to allow a low cost, high reliability, low latency and high bandwidth interconnect network within the BBU pool. There, open platforms and real-time virtualisation technology rooted in cloud computing are employed to achieve dynamic shared resource allocation and support of multi-vendor, multi-technology environments.

Figure 3 provides a visual representation of the functional split between BBU (on the left) and RRH (on the right, in green) functions.

The Cloud RAN enables large scale centralized deployments: it allows hundreds of thousands of remote RRHs connect to a single centralized BBU pool. In this scenario, latency limits indicated in the standards play a major role in the design and dimensioning of the system and in the degrees of freedom in positioning the BBU pools. For this reason, the maximum distance can be 20 km using an optical fiber link for 4G (LTE/LTE-A) system, but longer distance (40 km 80 km) is possible for 3G (WCDMA/TD-SCDMA) and 2G (GSM/CDMA) systems. As an example of actual deployments, some Asia operators claim to have deployments of C-RAN systems with 1200 RRHs centralized to one central office.

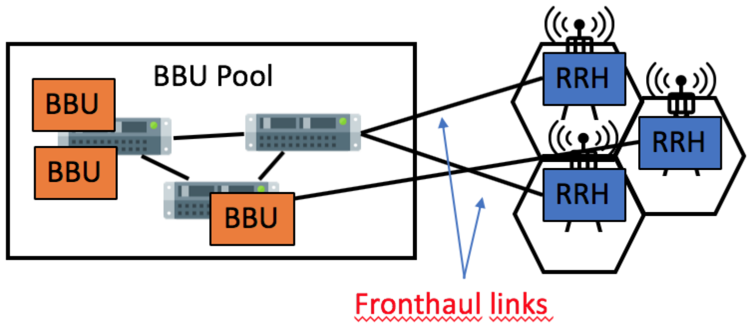

Fig. 2 The Cloud RAN Architecture concept: BBU is detached from RRH and move into a BBU pool.

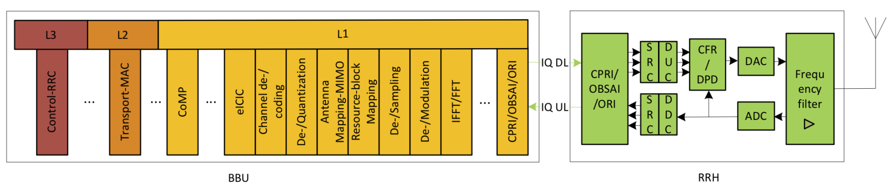

Fig. 3 Functional splitting between BBU and RRH. BBU and RRH communicate through the CPRI interface.

Another advantage of Cloud RAN is the native support to collaborative radio technologies. Indeed, using the C-RAN architecture, any BBU can talk with another BBU within the BBU pool with very high bandwidth (10Gbit/s and above) and extremely low latency (10us level). This is enabled by the interconnection of BBUs within the pool.

It should be noted that Cloud RAN is different from BBU Hoteling, or Base Station Hoteling. While in Cloud RAN BBUs can also be interconnected among them and flexibly re-located (as they practically Virtual Machines or containers), in hoteling the BBUs of different base stations are simply stacked together and no direct link among them is available.

Recently, a mathematical framework was developed at the CNIT Research Unit in Trento to study the potential benefits of Cloud RAN solutions in the framework of 5G. By using the model developed in [2], the team demonstrated the flexibility of the Cloud RAN architecture to adapt to the daily variations of the traffic pattern of a typical mobile operator. Moreover, they studied the possibility to consolidate the BBU VMs in order to provide further resource savings. Figure 4 provides an example of the achieved results, where the energy gain is analyzed for different BS configurations and at different hours of the day.

Please check [1] for further details on Cloud RAN and [2] for more details on the mathematical model.

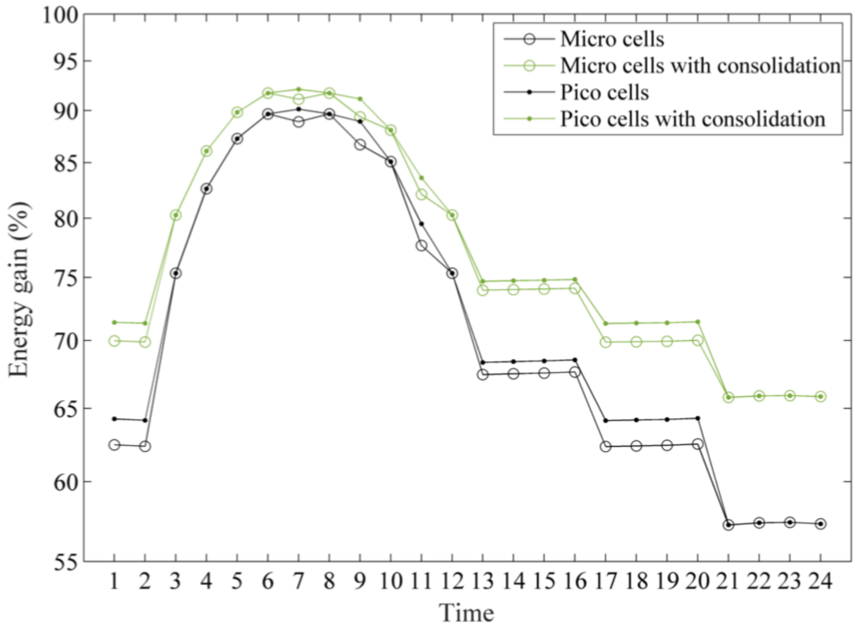

Fig. 4 Potential energy gains deriving from the deployment of the Cloud RAN paradigm: in presence of low traffic, fewer BBUs are required.

3 Mobile Edge Computing

Multi-access Edge Computing (MEC), or Mobile Edge Computing, is a network architecture concept that enables cloud computing capabilities and an IT service environment at the edge of the cellular network and, more in general at the edge of any network. The basic idea behind MEC is that by running applications and performing related processing tasks closer to the cellular customer, network congestion and latency are reduced and as a consequence applications can provide better performance. This is especially true in case of applications requiring low latency from the network, such as automated vehicle driving, remote operation of devices, etc.

MEC technology is designed to be implemented at the cellular base stations or other edge nodes, and enables flexible and rapid deployment of new applications and services for customers. Nevertheless, in order to deploy MEC solutions in an effective manner, cellular operators are required to open their radio access network (RAN) to authorized third-parties, such as application developers and content providers. This represents an additional requirement in the design of the 5G architecture.

The 5G system architecture specified by 3GPP and described in [3] is designed to address a wide set of use cases, which can be typically clustered into three groups:

Enhanced Mobile BroadBand (eMBB), massive Machine Type Communications (MTC), and Ultra-Reliable Low Latency Communications (URLLC).

As discussed on the introductory section, supporting all use cases with a common architecture has required significant changes in design philosophies both for the RAN and the core network. In the 5G system specification there are two options available for the architecture: one with the traditional reference point and interface approach, which represents an evolution of 4G LTE standard IP architecture, and the other where the core network functions interact with each other using a Service Based Architecture (SBA). Indeed, the SBA represents a big step forward in the virtualization and softwarization of the architecture.

Details on the SBA option of the 5G system architecture are provided in a white paper by ETSI [4]. In brief, the SBA framework is built around functions, that consume services and/or produce services. Any network function can offer one or more services. The SBA framework provides the necessary functionality to authenticate the consumer and to authorize its service requests, as well as flexible procedures to efficiently expose and consume services. For simple service or information requests, a request-response model can be used. For any long-lived processes, the framework also supports a subscribe-notify model.

ETSI Industry Specification Group on Multi-access Edge Computing (ETSI ISG MEC) defined an API framework aligned with the above principles. The functionality needed for efficient use of the services includes registration, service discovery, availability notifications, de-registration and authentication and authorization. All this functionalities are the same in both the SBA and the MEC API frameworks.

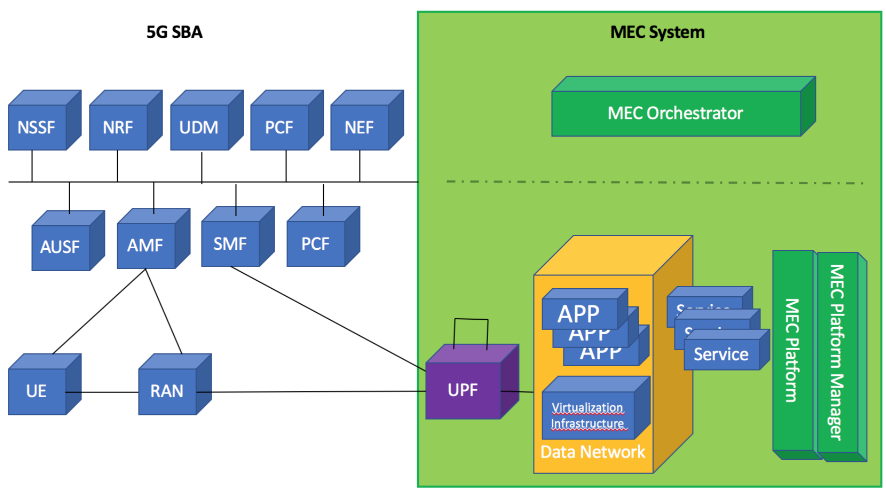

Fig. 5 MEC deployment in the 5G architecture proposed by ETSI.

The network functions and the services they produce are registered in a Network Resource Function (NRF), while in MEC the services produced by the MEC applications are registered in the service registry of the MEC platform. 5G Network Exposure Function (NEF) acts as a centralized point for service exposure and also has a key role in authorizing all access requests originating from outside of the system.

One of the key concepts in 5G is Network Slicing, that allows the allocation of the required resources from the available network functions to different services or to tenants that are using the services. The Network Slice Selection Function (NSSF) is the function that assists in the selection of suitable network slice instances for users, and in the allocation of the necessary Access Management Functions (AMF). A MEC application, i.e. an application hosted in the distributed cloud of a MEC system, can belong to one or more network slices that have been configured in the 5G core network.

The Unified Data Management (UDM) function is responsible for generating the 3GPP AKA authentication credentials, handling user identification related information, managing access authorization (e.g. roaming restrictions), registering the user serving NFs (serving AMF, Session Management Function (SMF)), supporting service continuity by keeping record of SMF/Data Network Name (DNN) assignments and performing subscription management procedures.

The User Plane Function (UPF) has a key role in an integrated MEC deployment in a 5G network. UPFs can be seen as a distributed and configurable data plane from the MEC system perspective. The control of that data plane, i.e. the traffic rules configuration, now follows the NEF-PCF-SMF route. Consequently, in some specific deployments the local UPF may even be part of the MEC implementation.

The resulting integrated architecture described in the white paper is presented in Figure 5, where the 3GPP 5G SBA system is shown on the left, while the ETSI MEC architecture is on the right.

The MEC system is presented on the right-hand side of Figure 5. The core module of the MEC system is the MEC orchestrator, a MEC system level functional entity that, acting as an AF, can interact with the Network Exposure Function (NEF). In some scenarios, the MEC orchestrator can also interact directly with the target 5G NFs. On the MEC host level it is the MEC platform that can interact with these 5G NFs, again in the role of an AF. The MEC host, i.e. the host level functional entities, are most often deployed in a data network in the 5G system.

Figures 6-9 depicts the available physical deployment alternatives for the deployment of MEC within 5G:

1. MEC and the local UPF co-located with the Base Station (Fig. 6).
2. MEC co-located with a transmission node, possibly with a local UPF (Fig. 7).
3. MEC and the local UPF co-located with a network aggregation point (Fig. 8).
4. MEC co-located with the Core Network functions (i.e. in the same data centre, Fig. 9).

Basically, the architecture enables to deploy MEC in different locations between the Base Station and a remote Data Center. Nevertheless, all deployments have in common the UPF that is used to steer the traffic towards the targeted MEC applications and towards the network. It should be noted that the MEC management

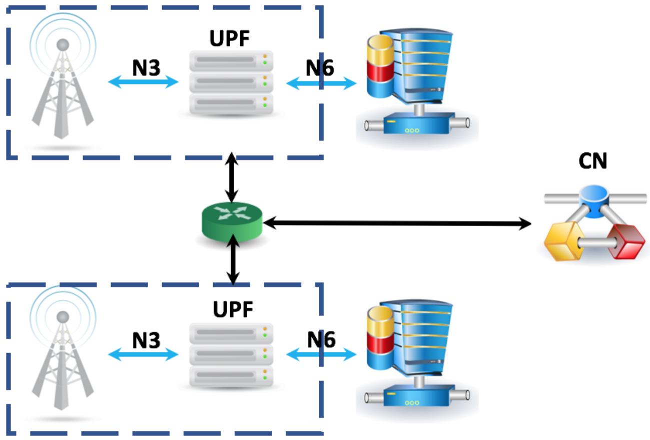

Fig. 6 Possible physical deployment of MEC in 5G networks - co-located with the Base Station.

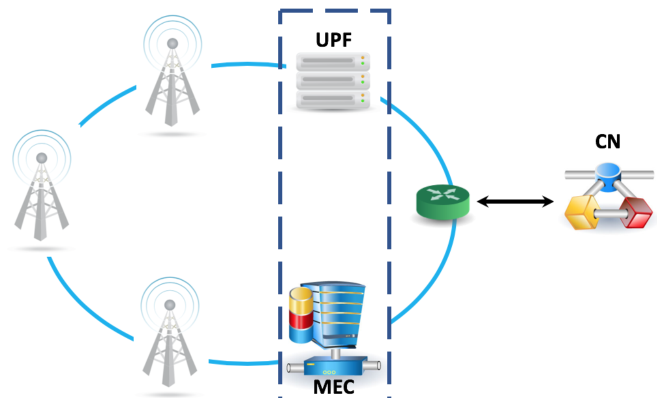

Fig. 7 Possible physical deployment of MEC in 5G networks - co-located with a transmission node.

system, which orchestrates the operation of MEC hosts and applications, may decide dynamically where to deploy the MEC applications.

4 Conclusions

Cloud RAN and MEC represent a big step forward in increasing the flexibility of the 5G network infrastructure, with the purpose to improve scalability and service support.

This chapter provided a brief introduction to the above concepts, providing some discussion on the potential benefits of such emerging paradigms and their integration in the upcoming 5G Standard.

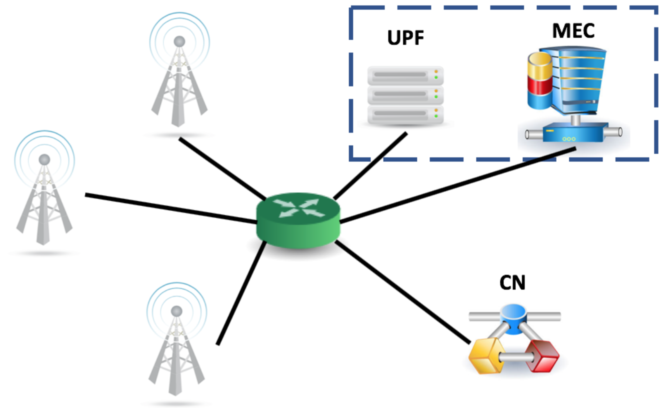

Fig. 8 Possible physical deployment of MEC in 5G networks - co-located with a network aggregation point.

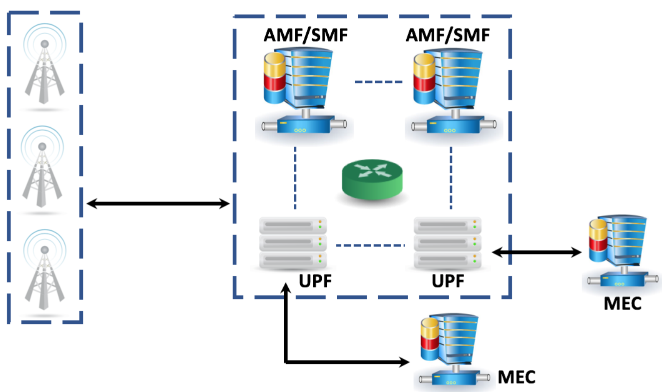

Fig. 9 Possible physical deployment of MEC in 5G networks - within the same data center.

References

1. A. Checko and H. L. Christiansen and Y. Yan and L. Scolari and G. Kardaras and M. S. Berger and L. Dittmann, "Cloud RAN for Mobile Networks—A Technology Overview", IEEE Communications Surveys Tutorials, Vol. 17, No. 1, pp. 405–426 (2015), doi: 10.1109/COMST.2014.2355255
2. R. Bassoli, M. Di Renzo and F. Granelli, "Analytical energy-efficient planning of 5G cloud radio access network," 2017 IEEE International Conference on Communications (ICC), Paris, pp. 1–4 (2017), doi: 10.1109/ICC.2017.7996871
3. 3GPP TS 23.501 V15.1.0, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2 (Release 15)" (2018-03).
4. ETSI White Paper No. 28, "MEC in 5G networks," First edition, June 2018, ISBN No. 979-10-92620-22-1