Providing Mission-Critical Services over 5G Radio Access Network

Rubén Solozabal, Aitor Sanchoyerto, Miren Cava, Bego Blanco, Hicham Khalifé, Mathieu Bouet, Damien Lavaux, Emmanouil Kafetzakis

Providing Mission-Critical services over 5G Radio Access Network

Rubén Solozabal[1], Aitor Sanchoyerto[1], Miren Cava[1], Bego Blanco[1], Hicham Khalife[2], Mathieu Bouet[2], Damien Lavaux[2] and Emmanouil Kafetzakis[3]

[1] University of the Basque Country (UPV/EHU), Pl. Ingeniero Torres Quevedo 1, Bilbao, Spain

[2] Thales Communications, 4 Avenue des Louvresses, 92230 Gennevilliers, France

[3] Orion Innovations, Aminokleous 43, 117 44, Athens, Greece

Abstract. 5G is called to introduce a major transformation in communication network architectures with its transition to cloud native networks. This transformation will enable new unique service capabilities that will drive the development of innovative applications. But, for 5G being successful in this task, the identification of the vertical sectors' requirements is of outmost importance in order to map them into the design of the network architecture. This paper proposes a novel Cloud-Enabled Radio Access Network (CE-RAN) architecture to support Public Safety services at the edge of the network. This proposal leverages Network Functions Virtualisation, Software Defined Networking and Mobile Edge Computing principles to provide Mission-Critical services through an isolated network slice. We suggest a CE-RAN architecture with two levels of cloudification to bring the service closer to the end user.

Keywords: Mobile Edge Computing, 5G, Cloud-Enabled Radio Access Network, CUPS, IOPS

1 Introduction

Even though 5G is obviously the natural evolution of current mobile broadband networks, it is called to make a disruptive change in communication network architectures, becoming the broader transition to cloud native networks. 5G is expected to bring new unique network and service capabilities, becoming a key enabler for emerging, revenue-generating 5G applications.

With this aim, the design of 5G is based on three main pillars: the flexibility of Network Functions Virtualisation (NFV), the programmability of Software Defined Networking (SDN) and the proximity to users of Mobile Edge Computing (MEC). The application of these trending technologies enables the creation of isolated service slices over the same physical network, allowing to adapt the network in order to meet the requirements of specific services.

All these forthcoming capabilities encourage the development of new applications and vertical industries. Consequently, it is also decisive for the success of 5G to identify the key vertical sectors' requirements and map them into the design of the network architecture. On this point, from Release 11 onwards 3GPP is considering the requirements of Mission-Critical (MC) communications as a central topic to address the key requirements of the next generation broadband Public Safety (PS) networks. There is a clear trend towards different forms of network sharing models as opposed to building out dedicated legacy PS networks.

In order to support the demanding set of requirements of the new verticals, 5G standardisation bodies aim to define a new Radio Access Network (RAN) architectural framework that leverages NFV and SDN to provide a complete virtualised ecosystem suitable for the execution of Virtual Network Functions (VNF). To this aim, we consider that the most appropriate place to build 5G infrastructure foundations is the edge of the network, in order to relieve the core network from the traffic that can be efficiently processed and served closer to the end user and, thus, to reduce the response time. We use Small Cell-as-a-Service (SCaaS) paradigm to build a highly flexible and scalable platform, able to support new business models and revenue streams by creating a neutral host market that reduces the operational costs by providing new opportunities for ownership, deployment, operation and amortisation.

The technical approach for exploiting the benefits of the centralisation of RAN functions is based on a two-tier architecture: a first distributed tier for providing low latency services and a second centralised tier with high processing power for access network control applications. Decoupling the control and user planes of the Radio Access Network (RAN) frees from the enormous fronthaul latency restrictions. The use of end-to-end network slicing mechanisms will allow sharing the infrastructure among multiple operators/vertical industries and customising its capabilities on a per-tenant basis.

This paper is organised as follows: Section 2 defines the use case for Public Safety applications, a vertical sector that is attracting appreciable interest in the latest years. Next, in Section 3 we describe the 5G Cloud-Enabled RAN with edge computing capabilities to support the defined used case. Then Section 4 proposes a solution that exploits MEC capabilities to provide MC services. Finally, Section 5 concludes this paper.

2 Description of the Public Safety use case

The current trends in the Public Safety ecosystem lead towards a future Mission-Critical (MC) communications framework based on standardised commercial technologies. In the recent years, several worldwide administrations have collaborated in the 3GPP to create the first operational solution in Release 13, named Mission Critical Push to Talk (MCPTT). This solution defines the service and network level specifications to provide mission-critical services over 4G LTE technologies. Currently, Release 14 evolved the solution towards a more mature state, including also other mission-critical services such as video and data. The adoption of the solution for future 5G networks will be specified during Release 15 normative work (scheduled for 2018).

In order to provide MC services the network provider must guarantee the required levels of QoS, security / privacy and resiliency.

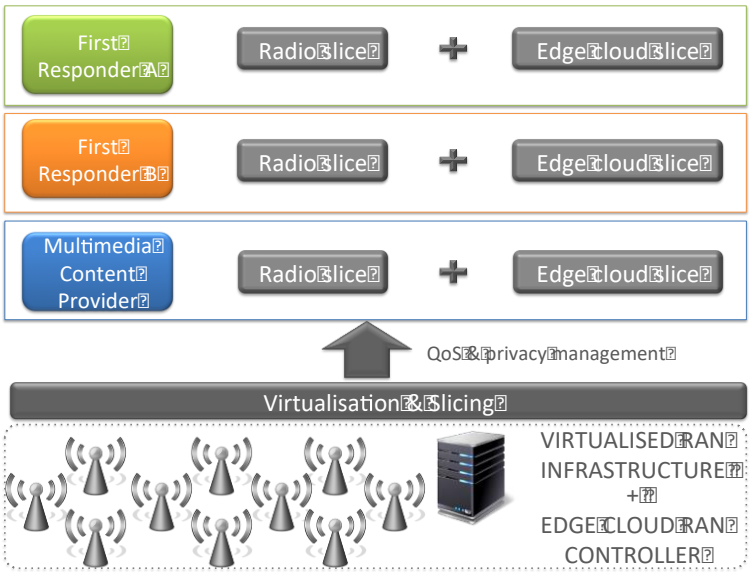

Fig. 1. Public Safety use case Stage 1 – Default service agreement.

This use case considers one or several Public Safety organisations using the resources provided by a third-party dense small cell deployment, which could be owned by the mobile network operator or by a venue owner such as the case of a stadium. In this case, the owner will exploit the proposed two-tier cloud edge platform to provide the required network slicing capabilities with dedicated characteristics to different types of tenants.

The public safety use case presented in this paper is organised in three main stages: **Stage 1**. In a first step, the small cell deployment owner is providing the required network slices to the different tenants (Fig. 1). Each network slice is composed by a volume of radio data rate over a coverage area (which is mapped by the edge cloud RAN Controller to a volume of small cell radio resources) and a volume of edge service capabilities (which is mapped to a volume of processing power in the lower level distributed cloud).

For the case of Public Safety organisations, normal operations may require a certain number of group communications supported in the area of the SC cluster. This requirement can be mapped to a number of radio KPIs in the SCs and the deployment of edge group communication service instances at the edge (i.e., MCPTT Application Servers deployed at the edge) to enhance the responsiveness of the service.

In addition to the QoS guarantees for each tenant, the deployment owner has to assure the required levels of isolation in the provisioning of the different network slices.

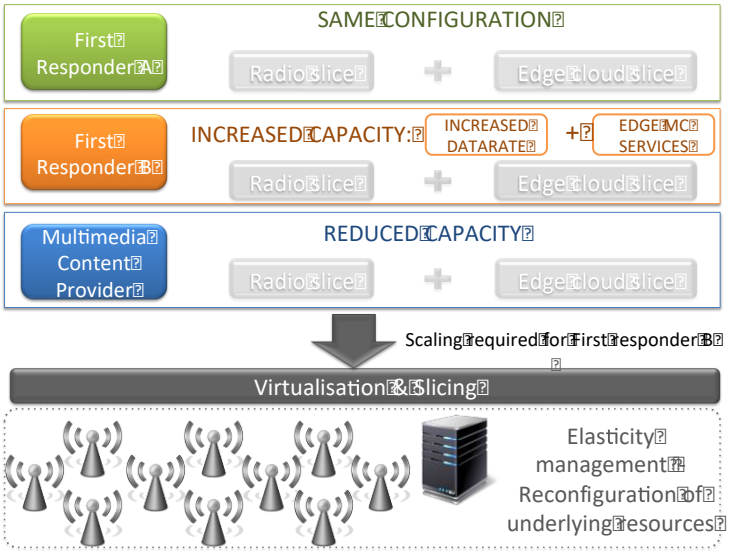

Fig. 2. Public Safety use case Stage 2 – Incident demanding extended capacity.

Stage 2. In case there is an incident in the area, the Orchestrator has to be able to react to the new service requirements (illustrated on Fig.2). A Public Safety organisation may require additional service capacity in the geographical area, in order to cope with increased number of first responders or additional types of service such as mission-critical video transmissions. Based on pre-arranged or on-demand service scaling policies, the Orchestrator would implement the new resource allocation scheme taking into account both radio (for the access connections) and cloud resources (for deploying more resource-consuming edge services).

The deployment of edge service instances serves a twofold objective: first, it enables close-to-zero delay in the mission-critical services; second, it allows maintaining the operability even when the backhaul connection is damaged.

Stage 3. Addresses the potential need for coverage extension (Fig. 3). The typical situation is that a Public Safety organisation decides to use a deployable system, e.g. to mitigate the damage of a macro base station. In the proposed use case, the deployable system also offers an (evolved) LTE connectivity to the first responders in the field, consolidating the interoperability requirements.

In order to better orchestrate the radio transmissions, the deployable system will be considered as a new small cell that can be dynamically added to the cluster. In this way, the enhanced SON and RRM features of the platform can be applied to the coverage extension unit. The connectivity of the deployable unit with the SC cluster would be made through a wireless backhauling technology.

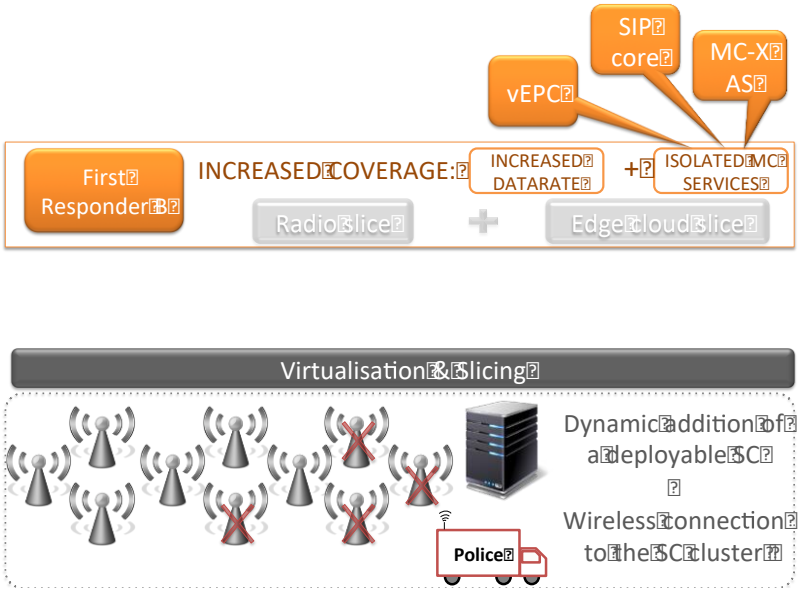

Fig. 3. Public Safety use case Stage 3 – Breakdown demanding extended coverage.

Once the Public Safety use case has been outlined, the next section describes the architecture proposed to provide PS services at the edge of the cloud-based RAN.

3 5G Cloud-enabled RAN with MEC capabilities

It is estimated that the user traffic volume over the mobile network will increase on the order of several magnitudes in the coming decade [1]. In addition, with the advent of Internet of Things (IoT) devices, outsourcing tasks to powerful clouds is becoming a trend due to the lack power of these devices. Nevertheless, traditional cloud services may end in a high-latency not compatible with real-time services. To overcome these problems, service providers can take advantage of MEC [2].

MEC allows bringing computation and network management to the edge of the mobile network, within the Radio Access Network (RAN), allowing service providers to deploy applications closer to the end-user. MEC reduces service latency as well as the bandwidth consumption since the core network is not or at least less involved in traffic between UEs and application servers. The relocation of the services on the edge of the mobile network will be of outmost importance in order to achieve technological expectations of 5G with an end-to-end latency of one milliseconds, over one million connections per square kilometre, and traffic rates ten times higher than in 4G.

Currently, MEC deployments are limited due to the lack of flexibility on the mobile network. In order to overcome these barriers, 5G will leverage NFV and SDN technologies to bring a cloud execution platform at the edge of the network [3].

For all these reasons, 5G will have to face architectural changes in the RAN. ENodeBs based on specific hardware will be aggregated in centralised cloud centres within the RAN, concentrating resources to simplify its deployment and management. RAN functions will be virtually executed and ideally connected to the remote radio heads through optical fibre fronthauls, as high fronthaul delays may degrade the performance of certain novel edge services that require close-to-zero latencies as prescribed by 5G objectives, As a consequence, this execution platform will be able to provide the computational power at the edge not only to virtualise network functionality but also third-party services.

The evolution of the current RAN to that envisioned in 5G represents a great effort for network operators. Therefore, intermediary solutions as [4] adopt the aforementioned principles to propose a novel distributed CE-RAN architecture evolving the traditional commercial Small Cells eNodeB (SCeNB) to Cloud Enabled Small Cells eNodeB (SCeNBce). These solutions rely on currently installed Physical Network Functions (PNF) and complement them along with a virtualised platform that supports the execution of innovative edge services as Virtual Network Functions (VNFs).

Proposals as [5] envision the coexistence of both centralised and edge cloud. This two-tier cloud-enabled RAN forms which is known as a hybrid cloud in which both centralised and edge clouds will cooperate to take advantage of the resulting peculiarities of the architecture, ending in the benefits described below.

The First-tier is the Edge Cloud. It remains distributed over a SCeNBce cluster for providing close-to-zero service latency directly from the networks' edge. The edge cloud is used to support the executions of VNFs as deep packet inspection, GTP encapsulation or distributed SON. The Edge Cloud will reduce the cost of high bandwidth fronthaul over long distances to centralised data centres while being able to provide close-to-zero latency services.

The edge cloud consists in a Multi-RAT 5G small cell cluster with its standard backhaul interface. It allows multiple core network operators to provide its services sharing the same physical infrastructure. In this environment, the small cell is the termination of the GTP-User Plane which communicates with each of the core networks.

The Second-tier is a high-scale cloud. The Central Cloud will provide high processing power for computing intensive network applications. It will benefit from its global vision of the underlaying infrastructure so as to be able to compute efficient scheduling algorithms. The central cloud hosts the RAN controller, which makes control plane decisions for all radio elements in the geographical area of the cluster, security, traffic engineering, mobility management, etc.

Both these clouds will be conceived as an integrated hybrid cloud infrastructure from the upper layers. In this context, a single orchestrator will coordinate a variety of both, being able to deploy VNFs along them in a unified manner.

3.1 Network slicing within the Radio Access Network

5G will enable logical network slicing, with which, a single physical network is partitioned into multiple virtual networks being able to personalize it for each type of

services. With this technology, slices can be optimized to guarantee multiple characteristics for specific services including latency or bandwidth requirements.

In order to provide an end-to-end service, network slicing is required not only in the operator's core network but also within the RAN. Service slices must be achieved over the same physical infrastructure. Therefore, computational and radio resources need to be shared and isolated between different services and tenants. In this context, Mission-Critical (MC) services must be provided over a prioritised slice, guaranteeing the access to the needed end infrastructure. As a result, resources are delivered in an elastic manner based on on-demand service and the scaling policies applied.

The envisioned neutral RAN is an infrastructure shared between multiple virtual network mobile operators and content service providers. Guaranteeing resource isolation in a cloud environment is the key element in order to obtain an architecture capable of hosting multiple virtual network operators and vertical industries providing their services using the same infrastructure.

3.2 Radio Resource Management within the Radio Access Network

Currently, radio resource allocation remains inadequate, especially in urban areas, where, due to the increased in traffic and limited spectrum, operators may have to decrease the size of the cells. The dense deployment of small cells leads to interferences among neighbouring base stations. In order to solve this, a separation between control from data plane of the Radio Resource Management (RRM) is proposed. This way, the radio controller could be allocated in the Central Cloud, benefiting from the overall view of the underlying architecture and making decisions to the whole distributed data plane on how to operate, and reducing interference in a centralised manner and increasing the spectrum utilisation efficiency.

Therefore, the RAN evolves not only to provide multi-operator radio access, but also to achieve an increase in the capacity and the performance.

3.3 Management and orchestration of edge services

The approach for network control centralisation does not only remain for radio interfaces, but also for service managing and orchestration. The management and orchestration (MANO) will operate at multiple SCeNBces clusters at multiple point-of-presence, transforming them in a uniform virtualised environment, enabling the creation of optimization algorithms that would act over the whole deployment. The main part is the Network Function Virtualization Orchestrator (NFVO). It composes service chains and manages the deployment of VNFs over the edge cloud. The NFVO delegates the management of VNF lifecycle to the VNFM, which oversees the instantiation, update and overall scaling and termination of VNFs.

Inside the MANO an important part is the Element Management System (EMS). It provides end-user functions for management of both PNFs and VNFs. It is responsible for set the management functionalities that other entities as the Network Management System of each operator will see.

The MANO also encompasses telemetry and analysis to capture relevant indicators of the network operation, since it is responsible for the SLA agreements.

4 Deploying Mission-Critical service on MEC infrastructure

Nowadays, 5G core architecture standardization is in an early stage [6]. It is envisioned that 5G New Radio will be first implemented using 4G EPC, give rise to what is known as 5G Non-Standalone. Therefore, the core nodes we are going to refer to belong to the legacy EPC core architecture.

The isolation that will be achieved in 5G enables the network to meet the specific requirements of services as Mission-Critical (MC) communications. As a reference, Mission-Critical Push-to-Talk (MCPTT) is an IP-based MC service that can benefit from the MEC infrastructure. This kind of service requires of a Session Initiation Protocol (SIP) core such as IP Multimedia Subsystem (IMS) to operate. An IMS aims to reach interoperability for session control in all-IP Next Generation Networks and currently are implemented as a centralized subsystem attached to the Evolved Packet Core (EPC) of each operator.

We propose to distribute MC services near the end-user, making use of MEC capabilities obtained throughout the virtualisation of the RAN. Distributing the User Plane (UP) in the proximities of the end-user benefits first responders from mouth-to-ear latency reduction. Thanks to NFV, it also enables service providers to scale the user plane on demand in specific locations straightforwardly.

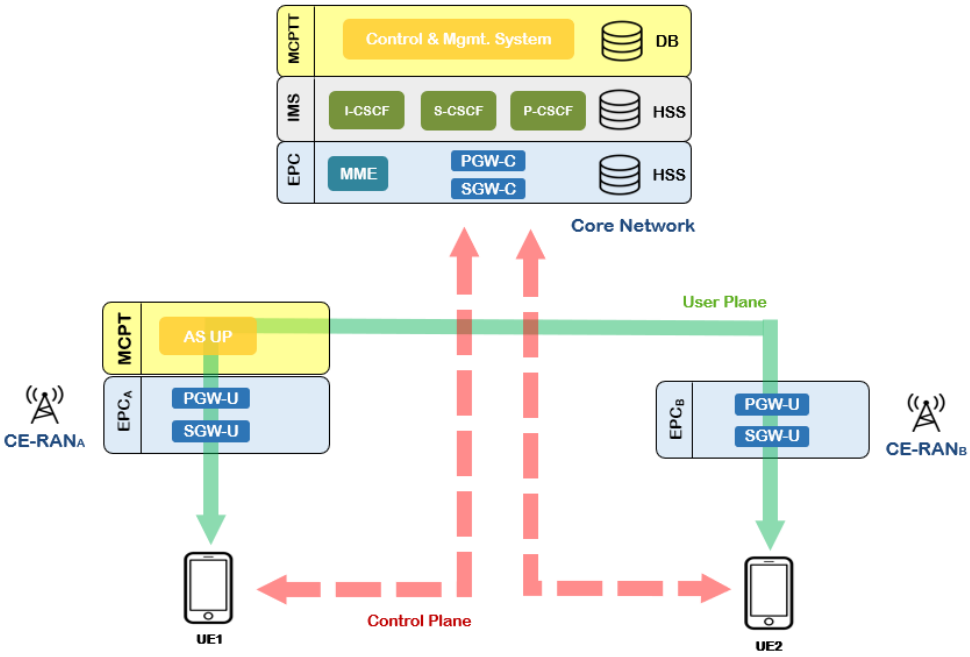

Fig. 4. Distributed MCPTT User Plane at the edge.

Distributing the service demands bringing the UP of the IMS and EPC to the edge of the mobile network. The deployment of an" IMS as a Service" (IMSaaS) and an" EPC as a Service (EPCaaS)" at the edge enables the necessary infrastructure to provide over-the-top services that otherwise would be located behind the operator's core network. As it will be discussed in the following section, it is common that none of the IMS nodes manage user data. And therefore, it will not be necessary to ensure with this layer near the user, service application located at the edge will be directly connected to the local EPC.

On the other hand, we propose to leave the Control Plane (CP) centralized due the complications that entails the synchronization of a distributed control.

4.1 Distributing core operator infrastructure at the edge of the network

Extending the current EPC infrastructure next to the edge allows reducing latency, enables scaling the service horizontally in specific locations and reduces traffic over the central core network. Nevertheless, the standardized EPC by 3GPP was never designed for a centralised network architecture, where the EPC nodes are part of the Core Network: Mobility Management Entity (MME), the Home Subscriber Server (HSS), the Serving Gateway (SGW) and the Packet Data Network Gateway (PGW). Recently, 3GPP has addressed an initiative called CUPS (Control and User Plane Separation) [7] with the aim to study possibilities for a separation of S/PGW functionality into user

plane functions and control plane functions, so that the user plane functions can be placed flexibly at the edge while the control plane functions could still remain centralized.

Local edge nodes are able to manage user data plane functions of SGW (SGW-U) and PGW (PGW-U) (illustrated on Fig. 4). Nevertheless, the replication of the EPC at the edge entails some limitations. Providing a stable data path to terminals changing their point of attachment to the network is the essential issue that will drive the new architecture design. In the legacy network architectures, a terminal's user traffic is always routed through a centralised node in the core network. This centralised node acts as an anchor point for the data path and ensures that IP packets reach the terminal irrespective of its point of attachment. Mobile networks need to adopt the distributed nature of routing IP providing mobile data path management on top of a distributed architecture because, in case of unmanaged IP mobility, the transmission could be disrupted. Nevertheless, there is a way to achieve application session continuity managed in upper functional layers. For example, a session that has been established using the SIP can survive IP address changes using the mobility management support built into those protocols.

In the same way that the EPC has been partially re-located at the edge, IMS shall be replicated as well. Nevertheless, 3GPP has not specified yet how the separation between the CP/UP of the IMS would be handled.

As mentioned before, an IMS is a signalling core, it's main operation is controlling. There are some nodes in the IMS architecture that manage UP (e.g. the Media Resource Function), yet these nodes are dispensable in a basic configuration. For this reason, the IMS do not appear as layer of the UP we are deploying at the edge. And service communication with IMS centralized nodes only will take place during control operations (e.g. during communication establishment).

Therefore, an EPCaaS at the edge entails the necessary infrastructure to provide over-the-top services with independent user plane without relying in the core network, being the appropriate host MC services as MCPTT.

4.2 Isolated E-UTRAN Operation for Public Safety

A specific scenario in which public safety communications can take advantage of MEC is in case of network failure. As defined in 3GPP technical specification 22.346 [8], isolated E-UTRAN aims to restore the service of an eNodeB or a set of interconnected eNodeBs without backhaul connectivity. The goal of Isolated E-UTRAN Operation for Public Safety (IOPS) is to maintain the maximum level of communication for first responders when the connectivity with the EPC is either unavailable or non-ideal.

In an emergency scenario, the execution platform located in the edge of the network could be used not only to bring the core UP near end-users but also to provide the CP, resulting in a fully capable local Core able to provide MC services within the isolated zone. In that direction, policies of backhaul disconnection must be agreed, as well as a methodology to change between a fully connected eNodeB to an isolated situation.

5 Conclusions

Considering the envisioned disruptive capabilities of forthcoming 5G communications and the highly demanding requirements of the services that will be supported, a radical RAN paradigm change is foreseen. It is also expected that this transformation will leverage NFV, SDN and MEC technologies as key enablers of this philosophy turnaround.

We rely on the two-tier Cloud-Enabled RAN as the platform to execute the virtualised services in end-user proximity. This architecture benefits the network operator from CAPex/OPex reduction, but also provides the necessary infrastructure to share the point-of-presence between multiple operators and third-party service providers.

We present the benefits of transferring the user data plane close to the end user, both in terms of QoS and from an operation prospective. To achieve this goal a separation between user and control planes throughout the vertical infrastructure is needed. In this paper, a use case of a Mission-Critical service on MEC is described. Enabling the deployment of user plane over at the edge of the mobile network benefits the service latency, but also facilitates the deployment and scaling on demand from the operators' perspective. In future work, we will analyse the impact of applying these MEC architecture principles over key performance indicators (KPI) for a Mission-Critical service.

We conclude highlighting the opportunities of the edge architecture in an emergency scenario. Under exceptional circumstances, in which the interconnections with the operators' core or even between base stations is damaged, a complete service deployment on the described architecture is able to maintain the service locally, at the expense of service restoration.

Acknowledgement

The research leading to these results has been supported by the EU funded H2020 5G-PPP project ESSENCE (Grant Agreement Nº 761592) and the Spanish Government's MINECO project 5GRANVIR (TEC2016-80090-C2-2-R).

References

1. S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," in *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36-43, May 2014. doi: 10.1109/MCOM.2014.6815891
2. T. X. Tran, A. Hajisami, P. Pandey and D. Pompili, "Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges," in *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54-61, April 2017. doi: 10.1109/MCOM.2017.1600863
3. Blanco, Bego & Fajardo, Jose Oscar & Giannoulakis, Ioannis & Kafetzakis, Emmanouil & Peng, Shuping & Pérez-Romero, Jordi & Trajkovska, Irena & Sayyad Khodashenas, Pouria & Goratti, Leonardo & Paolino, Michele & Sfakianakis, Evangelos. (2017). Technology pillars in the architecture of future 5G mobile networks: NFV, MEC and SDN. Computer Standards & Interfaces. 54. 10.1016/j.csi.2016.12.007.

4. Fajardo, Jose Oscar, et al. (2016). Introducing Mobile Edge Computing Capabilities through Distributed 5G Cloud Enabled Small Cells. Mobile Networks and Applications. 10.1007/s11036-016-0752-2.
5. Y. J. Ku et al., "5G Radio Access Network Design with the Fog Paradigm: Confluence of Communications and Computing," in IEEE Communications Magazine, vol. 55, no. 4, pp. 46-52, April 2017. doi: 10.1109/MCOM.2017.1600893
6. 3GPP, “Technical Specification Group Services and System Aspects; System Architecture for the 5G System,” 3rd Generation Partnership Project (3GPP), TR 23.501, 06 2017
7. 3GPP, “Architecture enhancements for control and user plane separations of EPC nodes” 3rd Generation Partnership Project (3GPP), TR 23.214, 06 2017
8. 3GPP, “Isolated E-UTRAN operation for public safety” 3rd Generation Partnership Project (3GPP), TS 22.346, 03 2017