IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 10, 2018, accepted June 7, 2018, date of publication June 20, 2018, date of current version July 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2849200

Exploitation of Mobile Edge Computing in 5G Distributed Mission-Critical Push-to-Talk Service Deployment

RUBÉN SOLOZABAL[1], AITOR SANCHOYERTO[1], ENEKO ATXUTEGI[2], BEGO BLANCO[3], JOSE OSCAR FAJARDO[1,2], AND FIDEL LIBERAL[1]

[1]Department of Communication Engineering, University of the Basque Country, 48013 Bilbao, Spain
[2]Nemergent Solutions SL, 48013 Bilbao, Spain
[3]Department of Computer Languages and Systems, University of the Basque Country, 48013 Bilbao, Spain

Corresponding author: Bego Blanco (begona.blanco@ehu.eus)

This work was supported in part by the European Commission under the H2020 5G-PPP Project ESSENCE under Grant 761592, in part by the European Union's Horizon 2020 Research and Innovation Programme (MONROE) through the Open Call MARiL Project under Grant 644399, and in part by Spanish MINECO through the 5RANVIR Project under Grant TEC2016-80090-C2-2-R.

ABSTRACT There is a growing interest in adapting mission critical public-safety communications from traditional private radio technologies toward 5G. The need for resiliency and strong delay requirements has resulted in mobile edge computing (MEC) emerging as a key technology to improve the deployment of public safety applications over the mobile network. This paper presents a non-standalone 5G ETSI MEC-based architecture for mission-critical push-to-talk (MCPTT) services. The proposal suggests a hierarchical distributed MCPTT architecture that allocates the user plane at the edge, keeping the control plane (CP) centralized for synchronization and assistance purposes. The MEC architecture enables the deployment of low latency services, due to the proximity of functional servers to the end-user, releases user equipment from heavy workloads, and benefits from horizontal scalability to provide dynamic allocation of network resources at specific locations. The proposed architecture is also beneficial in an isolated E-UTRAN operation situation, where in the case of backhaul connection loss, local MCPTT services can be straightforwardly deployed within an isolated group of eNodeBs.

INDEX TERMS Mobile edge computing, MCPTT, virtualized 5G, network function virtualization, IOPS, CUPS.

I. INTRODUCTION

Public authorities have assessed the benefits of integrating Mission-Critical (MC) Public-Safety (PS) Communications towards commercial mobile broadband standards [1]. For that purpose, Long Term Evolution (LTE) has become the reference technology for mission-critical communications. LTE and LTE-Advance (LTE-A) have been largely deployed and adopted as part of the 4G technology; nevertheless, they were not designed to comply with reliability, confidentiality and security standards required in mission-critical services. Therefore, the 3rd Generation Partnership Project (3GPP) has encouraged the evolution of LTE specifications to address these requirements. The first document dedicated to PS was launched in 3GPP Release 11. Since then, increased effort has been focused on PS challenges (illustrated in Fig. 1).

In parallel to the work focused on the evolution of PS and seeking for the improvement of the mobile services in terms of capabilities, scalability and deployability, standardization bodies are defining and targeting next 5G network harmonization and technology selection. This will soon be a reality with the upcoming releases of 3GPP. 5G Standard is expected to be extensively influenced by the MC solutions developed over LTE's all-IP core network. Along these lines, the main

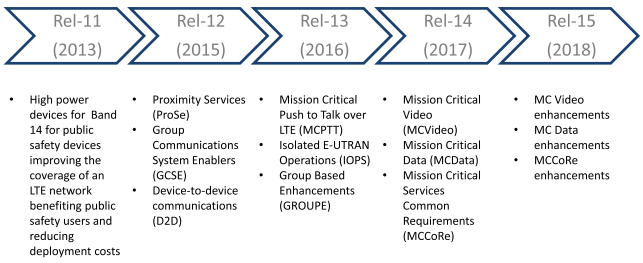

FIGURE 1. 3GPP Public Safety roadmap.

mobile operators have already deployed an IP Multimedia Subsystem (IMS) as the core framework in order to support and provide IP services such as Voice over LTE (VoLTE), or in the near future MCPTT (Mission-Critical Push-to-Talk) [2].

The evolved capabilities of 5G must meet the future growing needs of network bandwidth, massive number of connected devices (Internet of Things -IoT-) and support ultra-reliable and low-latency communications (URLLC). Additionally, there is a growing need for outsourcing computation to remote resourceful clouds, so that they could alleviate power and energy consumption in the end-devices. Nonetheless, computational offloading to the cloud may involve high latencies as well as the addition of even more load on the LTE core network. To overcome the aforementioned problems, services can take advantage of Mobile Edge Computing (MEC) capabilities. MEC allows bringing certain capabilities to the edge of the mobile network (e.g. computing, network management) within the Radio Access Network (RAN) [3]. This allows operators to run core services closer to the end-device and enables application developers and content providers to serve and adjust context-aware services using real-time radio access network information, user location, etc. MEC increases service responsiveness and reduces bandwidth consumption since the core network is not involved in the traffic between UEs and the application servers. The relocation of the service provisioning and the processing capabilities on the edge will be of outmost importance in order to achieve technological expectations of 5G such as well-known 1 millisecond latency, over one million connections per square kilometer, and traffic rates ten times higher than 4G. However, current MEC deployments are limited due to the lack of flexibility of the current network based on specific-purpose hardware. In order to overcome these barriers, 5G will leverage Network Function Virtualization (NFV) and Software Defined Networks (SDN) technologies [4]. These enhancements will bring the required softwarization, allowing MEC deployments to better fill the upcoming needs of network operators and their users [5].

In this paper we argue how MCPTT services could benefit from being deployed in an ETSI MEC [6] architecture. We describe the evolution the RAN will experiment in order to become a virtualized infrastructure capable of not only virtualize its native functionality but also providing scalability and deployability to services at the edge of the mobile network. It is important to underline that MCPTT services have architectural dependencies and require support from IP Multimedia Subsystem (IMS) and Evolved Packet Core (EPC) functionalities, being necessary to be built on top of the aforementioned layers. Therefore, any MEC-based MCPTT solution needs to consider the topology and signaling of each layer in order to have a complete top-down functionality. In this regard, some works have proposed MEC-based solutions than mainly focus on different alternatives related to the EPC [7]. However, to the best of our knowledge, there is no MEC-based proposal that has considered to bring the required functionality of IMS and MCPTT nodes to edge of the mobile network.

Therefore, this paper covers this gap by considering all three EPC, IMS and MCPTT layers and thus, presenting a complete MEC-based 5G architecture for MCPTT communications. Considering the constraints and architectural dependencies, we suggest the deployment of a distributed topology that allows fast deployment and increased responsiveness together with centralised control for synchronization purposes. Our proposal is able to not only improve the end-to-end transmission delay and QoS, but also to respond to emergency circumstances when the radio backhaul is down and a group of eNodeBs are in an isolated network situation (Isolated E-UTRAN Operation -IOPS-).

The paper is organized as follows. First, Section II deals with MEC in 5G. Next, Section III analyzes the proposed distributed architecture to achieve a distributed MCPTT service based on MEC capabilities. Section IV describes MCPTT as a locally deployable service to deal with an isolated situation. Both Section III and Section IV individually present the numerical estimation of the obtained benefit with the described proposal. Section V introduces MEC as a driver to Standalone 5G adoption. Finally, Section VI discusses the strengths and weaknesses of the proposed architecture and Section VII summarizes the main conclusions.

II. MEC IN 5G: THE EVOLUTION OF CLOUD-ENABLED RAN

In order to achieve a feasible MEC-based solution meeting all 5G requirements, a substantial change on the network paradigm becomes essential, specially at the network edge. 5G proposes to replace specific equipment in current eNodeBs with flexible centralised cloud centers within the RAN [8]. Centralised RAN systems will concentrate processing resources together in shared data centers simplifying deployment and management of the network, while improving CAPital EXpenditures (CAPEX)/ OPerating EXpense (OPEX). The explanation of the evolution of cloud-enabled RAN is divided into two subsections: a) The architectural evolution, and b) The standardized elements and the framework that all pieces form together.

A. ARCHITECTURAL EVOLUTION OF CLOUD-ENABLED RAN

The rise of Cloud Computing, along with SDN and NFV emerging technologies, makes possible to implement a Cloud Enabled RAN (CE-RAN) [9] that is able to provide computational power at the edge of the network, not only to network operators but also to third-party service providers (see MEC I in Fig. 2).

The evolution of the current RAN to the one suggested by 5G represents a great effort for network operators, particularly in terms of high-speed and low-latency fronthauls. Therefore, intermediary solutions such as [10] adopt the aforementioned principles to propose a novel distributed

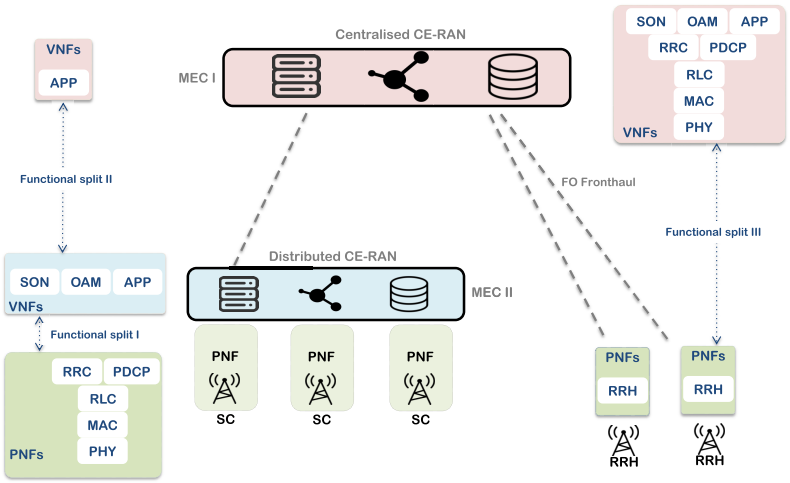

FIGURE 2. Hybrid 5G CE-RAN architecture.

CE-RAN architecture that fosters the adaptation of the current deployments towards 5G. This approach evolves the traditional commercial Small Cells eNodeBs to cloud enable ones. It relies on currently installed Physical Network Functions (PNF) complemented with a virtualized platform that supports the execution of innovative edge services as Virtual Network Functions (VNFs) (see MEC II in Fig. 2).

Hybrid approaches [11] will address the evolution from hardware-specific 4G platforms to emerging software-based 5G architectures. These proposals envision the coexistence of both centralised and distributed clouds cooperating in a coordinated manner, creating what is known as a hybrid cloud. According to this model, the execution of VNFs can be spread along the edge of the network and the centralised CE-RAN. This paper is aligned with the idea of having different MEC execution platforms cooperating in a hierarchical manner.

B. STANDARDIZED ETSI MEC ARCHITECTURE

Standardization bodies are devoted to address MEC specifications, following the advances in the cloud framework. Specifically, ETSI has put a great effort in standardizing the ETSI MEC [6] architecture. ETSI MEC is a framework that describes a mobile edge system that enables Mobile Edge Applications (MEAs) to run efficiently and seamlessly in a mobile network thanks to NFV & SDN technologies. ETSI MEC platform and MEAs can be deployed as VNFs in an ETSI NFV [12] infrastructure, those two technologies can blend together as described in ETSI GR MEC 017 [13]. ETSI MEC defines its own Mobile Edge platform Manager (MEPM) and Orchestrator (MEAO) that collaborate with the NFV Management & Orchestration (NFV MANO) [14] of the general-purpose cloud environment to operate MEAs. In this environment, SDN is in charge of intercepting traffic and forwarding it between VNFs. SDN can be employed to control and manage the network that interconnects distributed MEC servers, establishing dynamic and on-demand network connectivity for chaining VNFs.

III. MCPTT OVER DISTRIBUTED 5G NETWORK

An ETSI MEC compliant platform is used in order to deploy an MCPTT service in a distributed, virtualized and scalable manner, that brings the service within the CE-RAN, closer to the end user. The main goal of this proposal is to improve the QoS of mission critical services, being latency the major quality indicator for real-time communications and more precisely the mouth-to-ear latency (labeled as the third Key Performance Indicator -KPI3- in 3GPP [15]). By distributing the service and avoiding the transit of user data traffic through the central EPC, the mouth-to-ear latency (typical KPI for MCPTT) is expected to be reduced, increasing the QoS and responsiveness.

The provision of a MCPTT service requires a Session Initiation Protocol (SIP) core such as IMS. An IMS aims to reach interoperability for session control in all-IP Next Generation Networks. In traditional deployments it is implemented as a centralised subsystem attached to the EPC of each operator.

In order to benefit the MCPTT service, we propose to distribute its User Plane (UP) over the edge of the mobile network, which in turn requires to bring the underlying UP of the IMS and EPC. 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 near to the end-user that otherwise would be located behind the core network of the operator.

This infrastructure could be also beneficial in case of an isolated emergency scenario. In that situation, MCPTT could be deployed completely at the edge as an standalone service. For that purpose, it is required to bring not only the user plane (UP) to the edge but also the control plane (CP). This concept will be extended in Section IV.

In a virtualized network paradigm, MCPTT services will share the same physical CE-RAN infrastructure with other services. Therefore, the radio interface and computational resources will be divided among different service slices [16]. Mission critical services must be provided over a prioritized

slice, guaranteeing the access to the required end-resources. In an emergency situation, according to the service scaling policies, additional resources may be requested to cope with the increased number of first responders and therefore, resources must be elastically available to this prioritized slice.

As it will be described in the following subsections, transferring the UP implies the replication of the logical part of the nodes responsible of managing user data traffic: local virtualized EPC-vEPC-, local virtualized IMS -vIMS-, and local MCPTT virtualized application servers-vMCPTT ASs-; leaving centralised the CP of the EPC, IMS and MCPTT systems to maintain a global vision of the network. This section includes three subsections in order to individually explain each layer and the distribution of nodes, together with a final subsection with results related to the potential latency improvement that the proposed solution will provide.

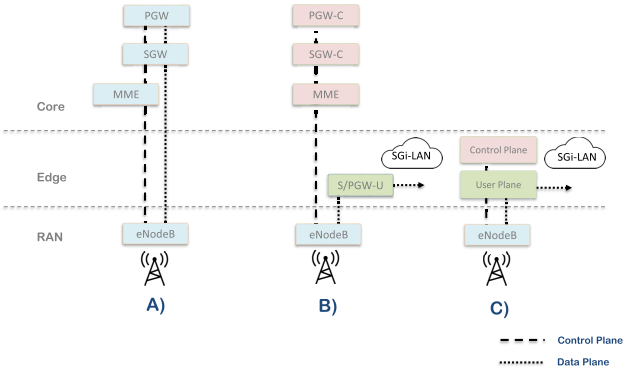

FIGURE 3. Evolution from centralised EPC to Control and User Plane Separation (CUPS).

A. DEPLOYMENT OF DISTRIBUTED EPC ON MEC ARCHITECTURE

In order to articulate a complete architectural description, the challenges of integrating a service in a distributed EPC are covered in this subsection. The virtualized EPC (vEPC) will run on top of a cloud infrastructure at the edge of the cellular network. By extending the current EPC infrastructure close to the edge, we are able to reduce mouth-to-ear latency, enable horizontal service scaling at specific locations and reduce traffic through the central core network. Nevertheless, the EPC standardized by 3GPP was originally designed for a centralised network architectures, where the EPC nodes are part of the Core Network, including the Mobility Management Entity (MME), the Home Subscriber Server (HSS), the Serving Gateway (SGW) and the Packet Data Network Gateway (PGW) (illustrated in Fig. 3A). Recently, 3GPP has addressed an initiative called CUPS (Control and User Plane Separation) [17] 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 flexibly placed (e.g. closer to the RAN) and independently scaled while the control plane functions could still remain centralised. As the 5G Core architecture (3GPP TS 23.501 [18]) is still in an standardization process, this initiative adds to the existing EPC the flexibility needed to meet 5G requirements. The network architecture composed by the 5G New Radio standard over an EPC is commonly referred as ''5G Non-Standalone'', and it is the one we are referring in the solution.

Following this concept, an evolution on the operators' core architecture is envisioned through the definition of an edge node with local EPC capabilities [7]. Edge nodes are able to manage user data plane functions of SGW (SGW-U) and PGW (PGW-U) (illustrated in Fig. 3B) and also if required, (e.g. in case of an isolated emergency scenario) edge nodes can act as a local controller entity on the edge composed by the control functions of the MME, SGW (SGW-C) and PGW (PGW-C) (illustrated in Fig. 3C).

However, the replication of the EPC at the edge entails some limitations that have been already discussed by the ETSI [19]. Those problems are related to the mobility management, session management, lawful interception, security, charging and UE identification. Some of them must be solved using a not-specified yet protocol between EPC and MEC, while others as the mobility management must be handled in upper protocol layers.

In standard handover procedures, the state of the user is transferred from current MME to the target one under the control of selected SGW, always under the supervision of the EPC and preserving at all time the PWG as the IP anchor point. This mechanism cannot be considered as a feasible straightforward mobility solution in a distributed deployment since the PGW-U may not be maintained during an ongoing communication. The provision of 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 traffic is always routed through a centralised node in the core network. This centralised node acts as an anchor for the data path and ensures that IP packets reach the terminal irrespective of its point of attachment. In this regard, mobile networks need to adopt the distributed nature of IP routing providing mobile data path management on top of a distributed architecture. In case of unmanaged IP mobility, the transmission could be disrupted. Nevertheless, there exist different mechanisms to manage application session continuity in upper functional layers. Concept that will be extended in Section III-C2.

B. DEPLOYMENT OF DISTRIBUTED IMS ON MEC ARCHITECTURE

Once described the EPC layer distribution to support our proposal, it is necessary to first describe the IMS layer to later be able to explain the MCPTT AS distribution. In the same way that the EPC has been partially re-located at the edge, the IMS UP needs to be replicated as well. Nevertheless, 3GPP has not specified yet how the separation between the CP & UP of the IMS would be handled.

Call state control function nodes (CSCF) as well as the Home Subscriber Server (HSS) are focused in control operations. These control nodes are necessary in an MCPTT communication, yet none of them route user data traffic. User data traffic might be routed through an IMS core, for example, in case of using a Media Gateway (MGW), however, in our simplified scenario user data traffic is directly forwarded to the corresponding application server. Therefore, communication with IMS core is only required during control operations (e.g. during communication establishment) and generally IMS nodes do not require to be distributed in the end user proximity.

C. DISTRIBUTED MCPTT SERVICE OVER IMS ON MEC ARCHITECTURE

On top of the distributed EPCaaS/IMSaaS infrastructure at the edge, different kind of services can be placed. In this paper, a solution to enable a close to the end user MCPTT service is proposed. For each MCPTT private and group call, there should be only one MCPTT server assuming the controlling role, while one or more MCPTT servers may be involved in participating role.

An MCPTT service may rely on other auxiliary servers in order to manage status information about the service such as Group Management Server (GMS), Configuration Management Server (CMS) or Identity Management Server (IdMS).

Continuing with the idea of having a centralised vision of the network accessed by local stateless servers, the management servers of MCPTT should remain attached to the central IMS, whereas MCPTT participating and controlling application servers in charge of the user data plane could be distributed on the edge (illustrated in Fig. 4).

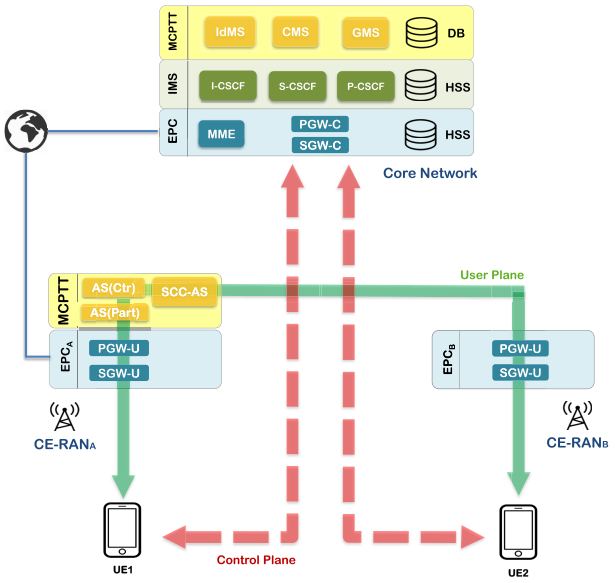

FIGURE 4. Distributed MCPTT service.

To raise the concept, some operating scenarios are described:

1) MCPTT PRIVATE CALL PROCEDURE

To illustrate this with a particular example, let us consider that UE1 is trying to establish a MCPTT communication with UE2 (as illustrated in Fig. 4). UE1 and UE2 are attached to the mobile network through CE-RAN$_A$ and CE-RAN$_B$ respectively. Both CE-RANs have a virtualized EPC deployed (vEPC$_A$ and vEPC$_B$) and therefore, users have data connectivity directly via their local PGW-U$_A$ and PGW-U$_B$. When UE1 initiates a MCPTT call with UE2, it requires from signaling managed by the centralised control plane of the IMS and MCPTT ASs to establish the communication. Once it is done, user data traffic is directly exchanged between PGW-U$_A$ and PGW-U$_B$ without crossing the core network.

The call establishment and control decision are taken centrally, being the IMS responsible for tracking user registration as well as for managing the logic of which edge server is optimum to manage a requested service. During the call establishment, the S-CSCF is responsible fixing the associated MCPTT AS to UE1. As there is a local MCPTT AS$_A$ deployed co-located with UE1, it is reasonable that it will be scheduled to manage the call.

In order to achieve a complete functionality a couple of underlying control procedures are needed: 1) the IMS should track user location in order to provide with the adequate local MCPTT AS; 2) the communication between EPC and IMS HSSs should exist to exchange user-related information.

2) MCPTT SERVICE CONTINUITY

During a handover on a ongoing transmission, MCPTT AS should remain irremovable as far as the communication is active in order to preserve session anchoring. An MCPTT session that has been established using SIP can survive IP address changes using the mobility management support built in this protocol [20]. In this case, a new element, the Service Centralisation and Continuity Application Server (SCC AS), is required in the architecture to act as a back-to-back user agent (B2BUA) within the IMS architecture. The SCC AS intercepts the communication an ensures end-to-end connectivity regardless the mobility of any of the end-points.

The procedure to maintain service continuity in a packet-switched to packet-switched network access transfer is described in in 3GPP TS 24.237 [21]. As mentioned, the call is anchored by the SCC-AS, being the media and if desired the signaling transferred from the IP source access leg to the IP target access leg seamlessly to the service. The SCC AS enables session control in the IMS home network for all voice sessions of end-nodes regardless of the network access type.

It operates at follows, after connecting to a new IP-CAN, obtaining a new IP address and discovering a Proxy Call Session Control Function (P-CSCF), the UE1 registers with the Serving Call Session Control Function (S-CSCF) over the new IP-CAN prior to initiating the packet-switched to packet-switched transfer procedure. Therefore, the IMS requires the UE1 to support simultaneous multiple registration and dual mode operation. Once the UE1 is connected with both interfaces it sends a SIP INVITE request message to the SCC-

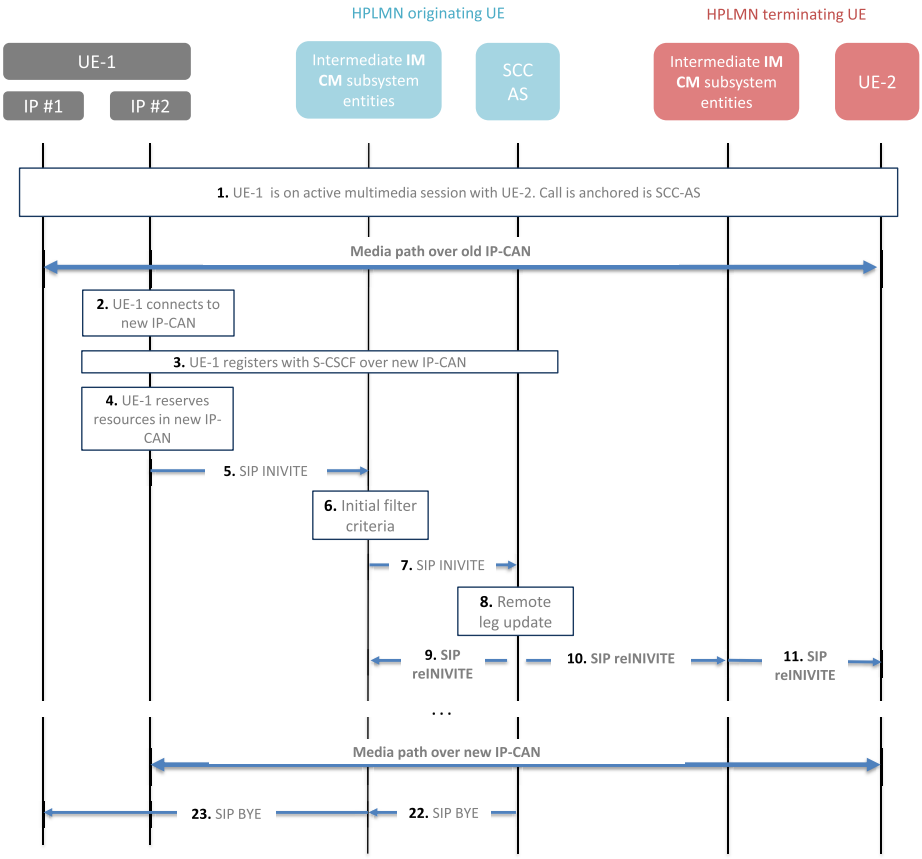

FIGURE 5. Service continuity in packet switched to packet switched network access transfer.

AS specifying in the Session Description Protocol (SDP) the new contact address that will be used for media over the new IP-CAN. The SCC-AS is responsible for sending a SIP reINVITE messages to the originating S-SCSF, which forwards it to the terminating network and finally to the UE2. A brief summary of the procedure in illustrated in Fig. 5.

During a ongoing call, UE1 traffic is redirected to S-CSCF$_A$ server even when it moves to another CE-RAN (hypothetical CE-RAN$_C$). When the call ends, a new session (re-)register would take place in the new CE-RAN$_C$ IMS location.

3) MCPTT GROUP CALL PROCEDURE

All MCPTT clients that belong to a single group are required to use the same MCPTT server for the entire group. The assigned MCPTT server responsible for carrying out the entire transmission, will be the one initiating the call and it should remain active throughout the conversation, even if the UE that initiated the call is not participating in the transmission anymore.

D. USER PLANE LATENCY RESULTS IN A DISTRIBUTED MCPTT SERVICE

After the top-down explanation of the three layers (EPC, IMS, MCPTT) that constitute our proposed MEC-based distributed MCPTT architecture, it is important to check the potential mouth-to-ear latency reduction that the user plane may well experience. To that end, we aimed at characterizing the current delays in the path within the UEs and MCPTT ASs to later on, deduce the portion of latency that would possibly be avoided due to the deployment of MCPTT ASs at the edge of the mobile network.

The presented results have been obtained using the MONROE project infrastructure [22]. MONROE is an EU project that provides a transnational open measurement platform for performance evaluations over cellular access network. It has a dedicated infrastructure with fixed and mobile nodes distributed over Norway, Sweden, Spain, Italy and Greece. It provides a flexible infrastructure to launch performance-related measurements and an accurate assessment of a variety of features in 4G cellular networks. The MONROE testbed has been utilized to perform latency measurements in the real-world over different commercial operators in different countries and network conditions. In addition to the provided nodes by MONROE, we have utilized a server in France with the necessary MCPTT ASs deployed. Fig. 6 shows the latency results with an Empirical Cumulative Distribution Function (ECDF) graph. The figure depicts three different values where the latency between the UE and the PGW of the operator is drawn in blue, the latency between

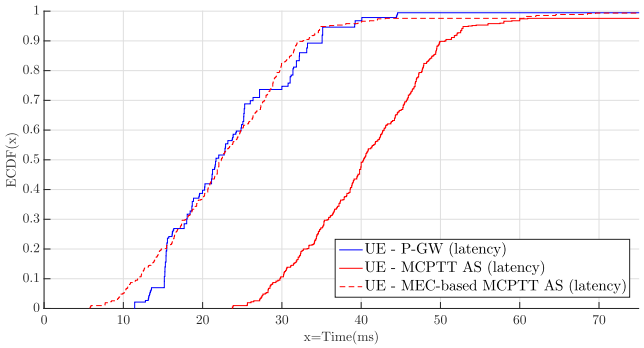

FIGURE 6. Experienced latencies in current mobile network and potential improvement of MEC-based solutions.

the UE and the conventional MCPTT server is shown in straight red line and the potential latency of the UE and the MEC-based MCPTT server is depicted in dashed red line.

After measuring the present path latencies in a wide variety of scenarios we have concluded that in a realistic situation over 90% of the communication have experienced end-to-end latencies close to 50ms (illustrated in Fig. 6 with a straight red line) to reach a centralised MCPTT server located behind a conventional EPC core. On the other hand, the latency experienced also in 90% of the tests between the end-user and the PGW has been about 30ms (see blue line in Fig. 6). Consequently, it is expected that latencies for the user plane of services deployed over MEC within the RAN will be reduced at least a 40% in each way. This is, considering that a transmission from one UE to other requires two hopes between the MEC-based server at the edge and the UE and taking into account highspeed connections between possibly different EPCs, the mouth-to-ear KPI would be greatly improved.

IV. ISOLATED E-UTRAN OPERATION FOR PUBLIC SAFETY

The efforts of the 3GPP to face public safety communications are focused on two directions: 1) enabling device-to-device communications for proximity-based services (ProSe); 2) continuity of the service in case of backhaul failure (IOPS).

As defined in 3GPP technical specification 22.346 [23], isolated E-UTRAN aims to restore the service of an eNodeB or a set of interconnected eNodeBs without addressing their backhaul connectivity [24]. The goal of IOPS is to maintain the maximum level of communication for PS users when eNodeB connectivity to the EPC is either unavailable or non-ideal. Isolated E-UTRAN can occur on top of nomadic eNodeB (3GPP TR 23.797 [25]) deployed on an emergency scenario or on top of fixed eNodeBs suffering failures.

The section is divided into two subsections: first we explain how the proposed distributed MCPTT services can appropriately respond to IOPS conditions and then we show numerical results of call setup times being improved.

A. LOCAL MCPTT SERVICES IN AN ISOLATED SCENARIO

In a crisis or tactical scenario, the described distributed MCPTT service can be fully locally deployed in an isolated eNodeB or accessible via a set of ones, in order to provide minimum PS services.

It is vital that field communications can be highly mobile and rapidly deployable to provide network access and coverage on scene. Currently, E-UTRAN is considered fixed, and the detection and discovery of a moving eNodeB remain unspecified. Enhanced eNodeBs must be able to discover other eNodeBs in their proximity, both directly or relying on the assistance of enhanced UEs. Inter-eNodeB connectivity links are not specified, therefore, this issue is left to vendor-specific solutions.

In current LTE architectures, eNodeBs are perceived as the active elements responsible for the management and control of the RAN. In contrast, UEs are passive clients from the eNodeB perspective, obeying certain rules and complying with the eNodeB's policies. However, UEs need to satisfy new requirements to take actively part in PS network. In addition to enable device-to-device communications, they would be able to extend PS network coverage acting as a relay node. Thus, extending eNodeBs coverage, informing to blind UEs of its accessibility to certain eNodeB and routing other UEs packages. But also, expanding UEs coverage vision in device-to-device communications [26].

We focus on enhanced features implemented on eNodeB in order to achieve IOPS requirements. Firstly, IOPS-capable eNodeB must be able to provide PS services as a standalone node. Secondly, it must be able to discover and deploy a mesh network between close proximity eNodeBs. Thus, nomadic or fixed eNodeBs that have lost connection can regenerate a local network to reach a node with locally deployed PS services. To assist in this interconnection, relay eNodeBs should exchange user traffic and signaling with other nodes belonging to the same Isolated E-UTRAN.

The initiation of Isolated E-UTRAN operation occurs when an eNodeB detects that there is an interruption on normal backhaul operation (S1 connectivity with the macro EPC). At this point the eNodeB can determine if it still has connectivity to other eNodeBs, reaching the macro EPC through other eNodeB's backhaul. In the case where it is not possible to reach it, local PS services must be deployed within E-UTRAN, resulting in Isolated E-UTRAN.

A minimal set of MCPTT capabilities need to be provided on the Isolated E-UTRAN. According to 3GPP TS 22.346 [23], Public Safety at least must be able to realize:

- Private MCPTT calls inside the Isolated E-UTRAN.
- Communication within pre-provisioned default groups based on organization membership.

Our proposal assumes that a local EPC, IMS, and MCPTT AS servers are deployed within Isolated E-UTRAN (illustrated in Fig. 7). As it has already been introduced in Section I, while no communication with the centralised EPC is available, both UP&CP shall be locally deployed. According to the proposal presented on the rest of the paper,

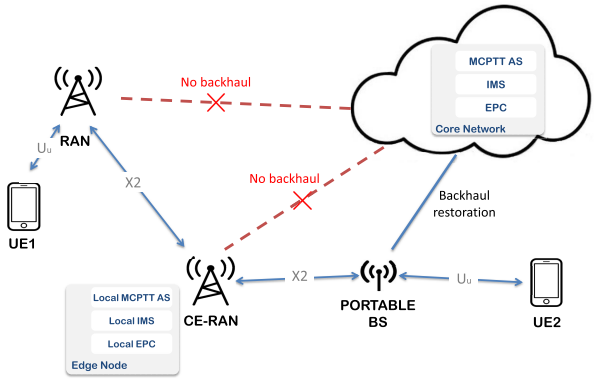

FIGURE 7. MCPTT service on isolated E-UTRAN operation.

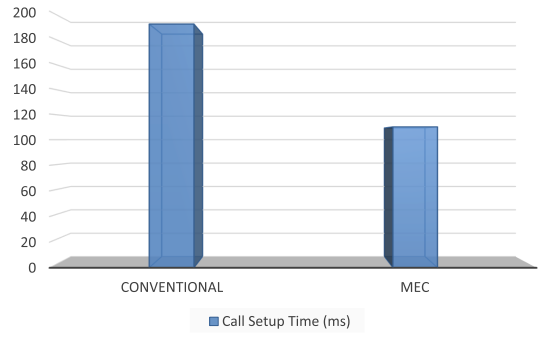

FIGURE 8. Improvement of call setup time in IOPS with MEC-based distributed MCPTT services.

no data-base information is stored locally. This is done to avoid challenges such as distributed synchronization or data integrity. When a local eNodeB changes from its normal operation to isolated mode, it can use cached HSS and MCPTT information updated before the loss of the backhaul or it can deal with pre-defined configuration. To restore normal operation, a deployable eNodeB can be placed in the emergency scenario. It can provision a backhaul connection via satellite or radio communication. In such case, the isolated E-UTRAN is able to restore connection with the macro EPC, returning to normal operation.

As demonstrated through the explanation, the deployment of MEC-based distributed architecture can also deal with critical use-cases of MC, providing this way a more complete and trustworthy option.

B. CALL SETUP RESULTS IN IOPS SCENARIO WITH DISTRIBUTED MCPTT SERVICES

After the explanation of how the MEC-based distributed MCPTT proposal fits and could be adjusted to precise MC use cases such as IOPS, it is important to numerically demonstrate the advantages that the proposal itself could bring. Besides the enhancement in the mouth-to-ear latency provided by the deployment of the UP on the edge, the deployment of the CP comes with other enhancement that is worth-mentioning.

Following the measurements to extract the portion of delay in each section of the end-to-end path, we performed MCPTT calls in order to characterize the total call setup time in current networks. By combining the call setup times and the extracted delays, we are able to deduce the potential reduction in call setup times when the CP remains on the edge. Fig. 8 shows the gathered average call setup times (labeled as ''Conventional'') and the potential evolution in a MEC-based MCPTT architecture (labeled as ''MEC''). The results depict that the call setup times could be improved from 190ms to 110ms due to the fact that the distributed CP avoids the transit of four messages (SIP INVITE message from the caller, SIP INVITE message to the callee, the call acceptance from the callee and the call acceptance to the caller) from the local node to the central one. In case of critical emergency conditions when the time is one of the most important assets, this enhancement allows faster call establishment.

V. MEC AS DRIVER TO STANDALONE 5G ADOPTION

One of the most notorious challenges on integrating MEC on a 4G EPC is the ability to perform traffic routing and steering between edge nodes. Several MEC deployments alternatives over the legacy network have been analysed in [19]. Nevertheless, there exist some architectural dependencies that limit its implementation. For this reason, SDN is envisioned to deliver user plane data more efficiently in a Standalone 5G system architecture.

The 5G Service Based Architecture (SBA) specified by 3GPP TS 23.501 [18] contains multiple control plane functional entities: Policy Control Function (PCF), the Session Management Function (SMF), the Application Function (AF), Access and Mobility Management Function (AMF), etc., and data plane functional entities like the User Plane Function (UPF). Integrating MEC data plane with the 5G system for routing traffic to the local data network and steering to an application is straightforward for this architecture.

The Application Function (AF) interacts with 5G control plane functions to influence traffic routing and steering. It acquires 5G network capability information and supports application instance mobility. In contrast to the current EPC, the 5G system is conceived to allow a more flexible deployment of the data plane, aiming to natively support edge computing. As a consequence, the MEC architecture can easily be integrated into that defined for 5G. Fig. 9 illustrates how MEC maps to the 5G system architecture, where the data plane defined in UPF elements is mapped to a MEC platform.

The MEC platform would perform traffic routing and steering function in the UPF. The PCF and the SMF can set the policy to influence the traffic routing in the UPF. Also, the AF via the PCF can influence the traffic routing and steering. Therefore, MEC in 5G is able to influence the UPF through the standardized control plane interface in SMF similarly to some of the EPC MEC deployment scenarios examined in 4G CUPS.

TABLE 1. Advantages and disadvantages of deploying a distributed service.

	Advantages	Disadvantages
Conventional	- *Ease of management*	- *Difficulty in scaling or deploying new services* - *Maintenance costs* - *Services are located away from end-users*
Distributed UP	- *Low-latency in user data traffic* - *Scaling ability*	- *Dependence on conventional architecture* - *Difficulty to orchestrate distributed services.*
Distributed UP&CP	- *Experience the benefits of a distributed UP* - *Ability to operate in standalone mode*	- *Partial view of the state of the service* - *Necessity of complex mechanisms to detect IOPS*

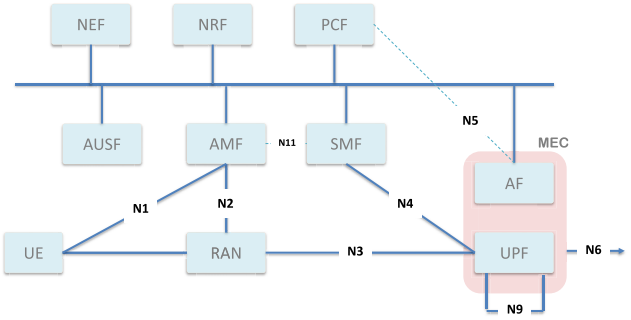

FIGURE 9. MEC mapping with 5G system architecture.

VI. DISCUSSION

To summarize and discuss the content of this paper we present a comparative table (view content in TABLE 1) gathering the advantages and disadvantages of the conventional network architecture and the distribution of MEC-based service planes.

One of the biggest advantages of the conventional architecture is that the management and orchestration is easier than with distributed nodes or logic. Nonetheless, there exists a great limitation in terms of service scalability and deployability. Since the equipment resides in specific hardware, deploying new services or scaling existing ones involves higher CAPex. Besides, due to the fact that the hardware is very specific, the associated maintenance costs are greater than with commodity hardware. In addition to the economic disadvantages, the services are located away from end-users, resulting in higher latency and therefore, worse KPIs and QoS.

In contrast, while deploying the UP of EPC, IMS and MCPTT at the edge, the overall latency of data traffic is significantly reduced, having a direct impact in the improvement of important KPIs such as mouth-to-ear latency in real-time MCPTT services. Moreover, the distributed architecture allows independently scaling up nodes at specific locations. Nevertheless, since the CP is kept centralised, the system experiences a continuous dependence on conventional architecture for controlling operations. Furthermore, distributed architectures entail a greater difficulty to appropriately orchestrate services at the edges in local nodes.

In case of a complete distribution of UP as well as CP at the edge, the experienced benefits would be the ones explained by the sole distribution of the UP plus the ability to operate in standalone mode in case of backhaul failure. However, as previously explained, since the logic of the state of the service remains centralised, the operative view at the edge is only partial or slightly outdated. In addition, there is a need for complex mechanisms that detect an IOPS scenario and are able to manage the reconfiguration to use local service nodes.

Considering the explained advantages and disadvantages, it is clear that each option provides certain positive aspects but also entails drawbacks. Our proposal covers both ones but considers that the effort requirement of a complete distribution of UP and CP is not worth as a generalized option. Despite this fact, the complete UP and CP distribution could be very relevant in drastic emergency conditions where the available deployment options are few or non-existent.

VII. CONCLUSION

The proposed Cloud-RAN in 5G architecture makes viable the extension of services near the end-user. This paper copes with the challenges of deploying a distributed MCPTT service taking advantage of MEC, where the service benefits from reduced latency and therefore improved KPIs. Additionally, the architecture facilitates deploying and horizontally scaling the service at specific locations.

The UP of MCPTT is distributed over an EPCaaS infrastructure at edge nodes, relying on centralised nodes allocated in macro nodes for controlling purposes. If needed, similarly any IMS node in charge of UP could also be deployed at the edge.

Besides, this paper has characterized the latencies present in the end-to-end path in current mobile networks for different operators and countries toward the same MCPTT AS. Considering the gathered latencies, the MEC-based MCPTT deployment would bring an about the 40% reduction in network latency, improving the mouth-to-ear KPI. In addition, only in IOPS conditions, if the CP is required to be deployed at the edge, the architecture would provide a significantly faster control and signaling, being able to reduce the call setup times from the centralised 190ms to the MC-based 110ms.

Several issues must be faced to meet the challenges of the proposed architecture. The main ones are the following: Firstly, a fully separation between UP & CP needs to be standardized for all protocol stacks required to provide the service. And secondly, ETSI MEC and ETSI NFV compliant VNFs need to be developed for the required service. As it has

been discussed, perform a MEC solution on the current EPC entails some architectural limitations. The 5G system is being conceived as a more flexible deployment of the user plane that facilitates the integration with the MEC architecture.

In relation with IOPS, politics of backhaul disconnection must be agreed, as well as a method to switch between a fully connected eNodeB to an isolated situation. And finally, a discovery protocol between eNodeBs is needed to dynamically add/remove nodes within the RAN, in normal operation as well as in an IOPS scenario to facilitate the recovery of the network.

ACKNOWLEDGMENT

Ruben Solozabal, Aitor Sanchoyerto, Bego Blanco, Jose Oscar Fajardo and Fidel Liberal were with the Networking, Quality and Security Research Group at the University of the Basque Country (UPV/EHU). (See http://det.bi.ehu.es/NQAS/). Jose Oscar Fajardo and Eneko Atxutegi were in Nemergent Solutions S.L. (See: https://nemergent-solutions.com/).

REFERENCES

[1] F. Liberal, J. O. Fajardo, C. Lumbreras, and W. Kampichler, "European NG112 crossroads: Toward a new emergency communications framework," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 132–138, Jan. 2017.

[2] *MCPTT Architecture and Flows*, document TS 23.379, 3rd Generation Partnership Project (3GPP), 2017.

[3] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 54–61, Apr. 2017.

[4] B. Blanco *et al.*, "Technology pillars in the architecture of future 5G mobile networks: NFV, MEC and SDN," *Comput. Standards Interfaces*, vol. 54, pp. 216–228, Nov. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0920548916302446

[5] O. Mäkinen, "Streaming at the edge: Local service concepts utilizing mobile edge computing," in *Proc. 9th IEEE Int. Conf. Next Generat. Mobile Appl. Services Technol.*, Sep. 2015, pp. 1–6.

[6] *Mobile Edge Computing (MEC); Framework and Reference Architecture, v1.1.1 Release*, document ETSI GS MEC 003, Mar. 2016.

[7] E. Cau *et al.*, "Efficient exploitation of mobile edge computing for virtualized 5G in EPC architectures," in *Proc. 4th IEEE Int. Conf. Mobile Cloud Comput., Services, Eng. (MobileCloud)*, Mar. 2016, pp. 100–109.

[8] P. Rost *et al.*, "Mobile network architecture evolution toward 5G," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 84–91, May 2016.

[9] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Netw.*, vol. 29, no. 1, pp. 35–41, Jan. 2015.

[10] J. O. Fajardo *et al.*, "Introducing mobile edge computing capabilities through distributed 5G cloud enabled small cells," *Mobile Netw. Appl.*, vol. 21, no. 4, pp. 564–574, Aug. 2016. [Online]. Available: http://dx.doi.org/10.1007/s11036-016-0752-2

[11] R. Solozabal *et al.*, "Design of virtual infrastructure manager with novel VNF placement features for edge clouds in 5G," in *Engineering Applications of Neural Networks*. Cham, Switzerland: Springer, 2017, pp. 669–679.

[12] ETSI. *Network Functions Virtualisation*. Accessed: Jul. 10, 2017. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/nfv

[13] *Mobile Edge Computing (MEC); Deployment of Mobile Edge Computing in an NFV Environment, v1.1.1 Release*, document ETSI GS MEC 017, Feb. 2018.

[14] ETSI. *Management and Organizationg*. Accessed: Jan. 9, 2018. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/nfv/open-source-mano

[15] *Universal Mobile Telelecommunications System (UMTS); LTE; Mission Critical Push to Talk (MCPTT) Over LTE; Stage 1, Version 13.3.0 Release 13*, document 3GPP TS 22.179, 3rd Generation Partnership Project (3GPP), Jan. 2016.

[16] R. Ferrus, O. Sallent, J. Perez-Romero, and R. Agusti, "On 5G radio access network slicing: Radio interface protocol features and configuration," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 184–192, May 2018.

[17] *Architecture Enhancements for Control and User Plane Separations of EPC Nodes*, document TR 23.214, 3rd Generation Partnership Project, Jun. 2017.

[18] *Technical Specification Group Services and System Aspects; 1. System Architecture for the 5G System; Stage 2, Version 15.0.0 Release 15*, document 3GPP TR 23.501, 3rd Generation Partnership Project (3GPP), Jan. 2018.

[19] F. Giust *et al.* "MEC deployments in 4G and evolution towards 5G," ETSI, Sophia Antipolis, France, White Paper no. 24, 2018.

[20] *MCPTT Service Continuity*, document TS 24.379, 3rd Generation Partnership Project (3GPP), 2017.

[21] *IP Multimedia (IM) Core Network (CN) Subsystem IP Multimedia Subsystem (IMS) Service Continuity*, document TS 24.237, 3rd Generation Partnership Project (3GPP), 2017.

[22] Ö Alay *et al.*, "Experience: An open platform for experimentation with commercial mobile broadband networks," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2017, pp. 70–78.

[23] *Isolated E-UTRAN Operation for Publc Safety*, document TS 22.346, 3rd Generation Partnership Project (3GPP), 2017.

[24] J. Oueis, V. Conan, D. Lavaux, R. Stanica, and F. Valois, "Overview of LTE isolated E-UTRAN operation for public safety," *IEEE Commun. Standards Mag.*, vol. 1, no. 2, pp. 98–105, Jul. 2017.

[25] *Study on Architecture Enhancements to Support Isolated EUTRAN Operation for Public Safety*, document TR 23.797, 3rd Generation Partnership Project (3GPP), 2015.

[26] R. Favraud, A. Apostolaras, N. Nikaein, and T. Korakis, "Toward moving public safety networks," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 14–20, Mar. 2016.

RUBÉN SOLOZABAL received the M.Sc. degree in telecommunications engineering from UPV/EHU University in 2015. He is currently pursuing the Ph.D. degree with particular interest in public safety over 5G architecture. He was with the R&D Division of Ingeteam S.A. for three years focusing on industrial communications.

AITOR SANCHOYERTO is currently pursuing the Ph.D. degree with UPV/EHU, with a focus on MC Services Management in 5G. He is currently an Engineer with over 20 years of experience in telecommunications projects in Public Safety Sector: Defense, National Security and Transport. He is specialized in interoperability protocols between organizations with different levels of security and radio technology, whose objective is to be able to establish voice and data communications in a safe way. In recent years, he has been actively working on convergence projects from narrow-band radio communications to broadband communications based on the 3GPP-defined standard for mission-critical services.

ENEKO ATXUTEGI received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of the Basque Country in 2011 and 2014, respectively, and the Ph.D. degree in telecommunications engineering from the University of the Basque Country in 2018, with a focus on analyzing the open-issues regarding the interaction of mobile networks and transport protocols and proposing cost-effective solutions to solve them. His research interests also include MEC and C-RAN topologies and NFV.

BEGO BLANCO received the B.S. and M.S. degrees in telecommunications engineering from the University of the Basque Country, Spain, in 2000, and the Ph.D. degree in telecommunications engineering from the University of the Basque Country in 2014. She is currently a Lecturer and a Researcher with the Faculty of Engineering, Bilbao. Her research interests include PQoS/QoE/QoS assessment and multicriteria optimization in 5G networks.

JOSE OSCAR FAJARDO received the Ph.D. degree in the area of adaptive management of mobile multimedia services in 4G and 5G networks from the Faculty of Engineering, University of the Basque Country, Bilbao, in 2016. He is a Co-Founder and the CEO with Nemergent Solutions SL, Bilbao, Spain, focused on the development, experimentation, and provisioning of 3GPP-based mission-critical services over 4G LTE networks. Within his responsibilities, he manages the company's private and public projects, including the R&D projects related to new mission-critical applications and new service deployment schemes, such as 5G and NFV. Since 2003, he has been a Research Fellow with the Faculty of Engineering, University of the Basque Country. He has co-authored over 50 journal and conference papers, mainly in areas of QoS/QoE and service performance assessment.

FIDEL LIBERAL is a well-recognized expert in the Mission Critical communications environment. He is currently with the University of the Basque Country, where he leads different mission-critical communications and 5G related research and development projects. Among them, he is currently the coordinator of the Mission Critical Open Platform project. He has co-authored more than 75 international journal and conference papers in different telecommunication areas, most of them in broadband mobile networks.

He has been very active in the field of Public Safety Communications related to NG911/NG112 systems and next-generation Public Safety systems. As a result of an intense dissemination activity, he is connected with some of the most relevant stakeholders in the sector such as EENA, PSCE, TCCA, NIST-PSCR and standardization bodies such as ETSI, 3GPP, and ITU-T. In 2017 he served as a technical expert in NG112 and MCPTT Plugtests, interoperability events organized by ETSI. He has also co-founded two security and mission-critical communications related Spin-Offs.