5G Framework for automated network adaptation in Mission Critical Services

Elisa Jimeno
ARI, Telecom Sector
Atos
Santander, Spain
elisa.jimeno@atos.net

Jordi Pérez-Romero,
Irene Vilà Muñoz
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
jorperez@tsc.upc.edu,
irene.vila.munoz@upc.edu

Begoña Blanco,
Aitor Sanchoyerto
University of the Basque Country, Bilbao, Spain
{begona.blanco,
aitor.sanchoyerto}@ehu.eus

Javier Fernández Hidalgo
Fundació i2CAT
Barcelona, Spain
javier.fernandez@i2cat.net

Abstract— ***Mission Critical Services (MCS) are gaining interest among network operators to offer alternative communications than conventional trunked radio systems. They promise a simplified management of cloud and radio resources for service deployment. However, the network capabilities should be adapted for the changing conditions, to assure low-latency and reliability for such applications. This paper presents an on-going work on utilising 5G technology for Mission Critical Push To Talk (MCPTT) services. It describes some design elements and evaluates 5G ESSENCE architecture that enable mission critical applications.***

Keywords— 5G, Mission Critical Applications, MCPTT, resource allocation, monitoring, Public Safety.

I. Introduction

The Fifth Generation (5G) network technology is envisioned to support diverse number of services and applications. Meeting the key performance requirements will be a big challenge in itself. “IMT for 2020 and beyond” is envisaged to develop a framework based on the recommendations made by the International Telecommunications Union (ITU) [1] focusing on three usage scenarios, with vastly heterogeneous requirements: eMBB (enhanced Mobile Broadband) focused on the increase demand of services such as AR(augmented reality)/VR (Virtual reality) for multimedia and content rich applications with higher data-bandwidth and moderate latency; URLLC (Ultra Reliable Low Latency Communications) for a robust and trustful communication; and mMTC (massive Machine Type Communications) characterized by the massive connectivity of IoT devices with low data transfer and low energy consumption.

In the context of the 5G ESSENCE project [2] several use cases have been identified to validate the Key Performance Indicators (KPIs) of 5G technology. One of these use cases deals with the provision of Mission Critical Services (MCS), imposing strict requirements associated to URLLC, which are not yet fulfilled by traditional communications systems.

Current Land Mobile Radio (LMR) systems, including conventional, trunked and secure communications, commonly used as push to talk wireless communication systems, rely on their own allocated spectrum for reliable communication, but they have several constraints such as narrowband channels that neither match the operational procedures of mission critical users, nor support interoperability between devices and collaboration from agencies and jurisdictions. Moreover, they are also expensive as they lack economies of scale. This situation poses a big limitation on the data transmission that neither allows rich content information nor multimedia data. In turn, while reducing the cost of the infrastructure, the 5G network will provide more reliable communication for first responders, enabling better device interoperability for the different actors, and applications services such as location, decision-making, access to more complete information, database, or video sharing. Apart from these features some other capabilities need to be offered by 5G when supporting MCS, including: High availability and reliability, prioritizing user’s services and traffic type, and restructuring in case of lack of infrastructure.

5G ESSENCE project also addresses the paradigm of Edge Cloud computing exploiting the benefits of the centralization of Small Cells functions to be offered as a service. In this way, the vertical industry of Public Safety can enhance Mission Critical applications by exploiting edge computing capabilities in a localized network deployment thanks to the ultra-lightweight, instantly deployable small cells, improving coverage, latency and reliability and at the same time changing the market perspective for operator’s revenue streams.

Within this context, mobile network operators (MNO) are able to supply power to the small cells (SC) faster and less expensively in places where power is not easily available, allowing MNOs to provide a wider coverage in unconnected areas, and the use of battery backups or generators at the centralized location to support busy or mission-critical SC.

This paper presents a practical solution towards bringing Mission Critical Applications towards the 5G technology. It describes some design ideas and concepts of an on-going work to enhance communications for public safety. The rest of the paper is organized as follows. Section 2 introduces the project background and the components used in this work. Section 3 describes the application to be demonstrated. Section 4 the scenario description and Section 5, 6, 7 presents the components involved and the configuration for the service deployment. Finally, Section 8 concludes the paper.

II. 5G Essence architecture

5G ESSENCE [2] project proposed a solution based on Multi-access Edge Computing (MEC) [3] capabilities, which supports delivery of high performance services with extremely low-latency and better service resilience due to the contextual information retrieved from the telemetry module.

Figure 1 depicts the high-level architecture of the 5G ESSENCE project. The cloud infrastructure uses a two tier architecture models: The Main Data Centre (DC), more powerful for intensive network applications (which require more computational resources); and the Light DC, closer to the user location, to provide low latency for services at the edge but more limited in computational resources. Moreover, the Light DC incorporates a Multi-Radio Access Technology (RAT) small cell that provides radio coverage through 5G, LTE and Wi-Fi technologies and enables radio resource sharing among different tenants. The combination of the small cell and the light DC is referred to as Cloud Enabled Small Cell (CESC). The 5G ESSENCE architecture allows the management of both tiers from the CESC manager, which includes an orchestration layer as well as Network Management System (NMS) and Element Management System (EMS) functions. Apart from that, it also includes monitoring and analytics capabilities to support the different decision-making processes.

The architecture also includes a centralized Software Defined-Radio Access Network (cSD-RAN) controller for decoupling the control and user plane of the Radio Access Network (RAN), supporting specific Radio Resource Management (RRM) and Self Organizing Network (SON) functions that control the operation of the CESCs of a cluster. This feature allows delivering high performance services with very low latency for critical services.

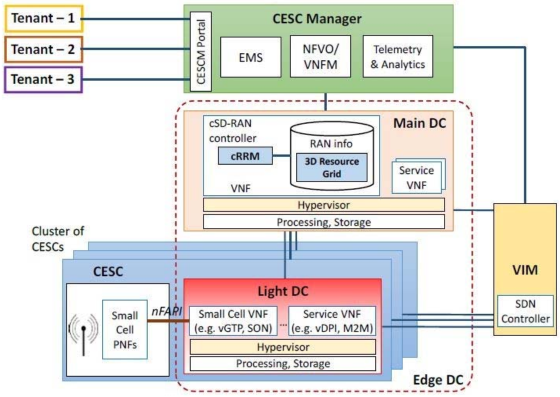

Figure 1: 5G ESSENCE High Level Architecture

III. MCPTT APPLICATION FOR FIRST RESPONDERS

Mission-critical push-to-talk (MCPTT) is a public safety mission-critical voice communication type, aimed at the coordination of emergency teams that are organized in groups [4]. It provides an arbitrated method by which two or more users may engage in communication. Users may request permission to transmit (e.g., traditionally by means of a press of a button) and the MCPTT service provides a deterministic mechanism to arbitrate between requests that are in contention (i.e., Floor control). When multiple requests occur, the determination of which user's request is accepted and which users' requests are rejected or queued is based upon a number of characteristics (including the respective priorities of the users in contention). Besides, the MCPTT service provides a means for a user with higher priority (e.g., MCPTT Emergency condition) to override (interrupt) the current talker. MCPTT Service also supports a mechanism to limit the time a user talks (hold the floor), thus, permitting users of the same or lower priority a chance to gain the floor. As it appears, the management of this type of half-duplex communications is not trivial, since it requires an appropriate management of priorities and privileges to allow communication.

The standardized MCPTT service imposes special requirements that include, among others, high availability and reliability, very low latency, support for one-to-one and group calls, talker identification and high audio quality for clear interchange of information. 5G ESSENCE common orchestration of radio, network and cloud resources is expected to significantly contribute to the fulfilment of the tight requirements of a MCPTT service (especially those related to latency), providing the tools to share both radio and edge computing capabilities between mission critical and commercial users.

The deployment of a complete MCPTT service includes the features that are described next:

- **Group calls:** instant one-to-many mobile voice communications that avoid the dial-ring-answer steps of a regular phone call. Typically, a group call allows only a user to speak at a time via half-duplex communication and provides call floor control mechanisms for managing the right to transmit at a point in time during an MCPTT call. To start a group call, the caller just selects the target group, presses the PTT button, speaks and the voice message is delivered instantly.

- **Private calls:** allow two MCPTT Users to communicate directly with each other without the use of MCPTT Groups (i.e., in a one-to-one manner). They leverage many of the functions and features of MCPTT Group Calls, such as MCPTT User identity and alias information, location information, encryption, privacy, priority, and administrative control.

- **Emergency calls:** pre-emptive calls due to an emergency condition. Upon the request of and emergency call, on-going calls can be terminated in order to free up resources for a higher priority call request.

After this brief description of the MCPTT service, the next section describes the proposed scenario for the validation of the service deployed on the 5G ESSENCE platform.

IV. SCENARIO DESCRIPTION

The validation of the complex communication system of Mission Critical applications, and the different capabilities they must offer, cannot be based on a static scenario, since one of the objectives to be proven is the elastic allocation of resources attending to different levels of emergency conditions detected by the monitoring system. We propose a deployment topology in three main stages, as shown in Figure 2.

At the beginning, in a situation under normal circumstances, the system instantiates the network slices that correspond to a default service agreement. Here, the first responder only needs a reduced amount of access capacity and communication features for its normal operations. Then, triggered by an emergency incident, the first responder requires increased capacity in terms of both data rate and edge computing resources, in order to serve a higher number of communications and/or public safety users. This situation may involve a deterioration of the service for legacy users, since their network slice must be reduced in order to appropriately allocate the higher priority MCPTT service. Finally, we will demonstrate how the service responds to an extreme situation of damaged infrastructure where a coverage extension is needed. In this situation, backhaul connectivity is lost, all the resources must be dedicated to the MCPTT network slice and the public safety organization may dynamically add new access points to the network in order to improve connectivity.

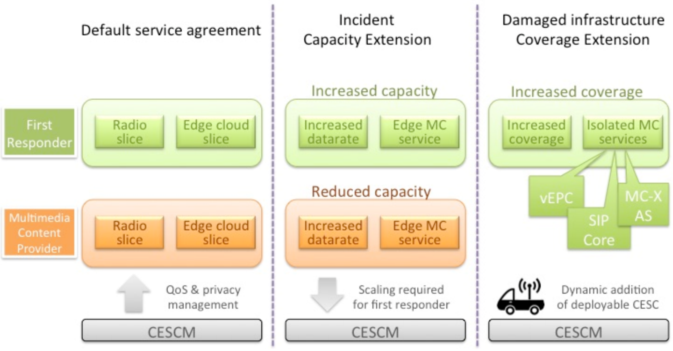

Figure 2:Deployment of the MCPTT validation scenario.

The challenge consists on transparently and elastically allocating the available cloud and radio resources to the variety of actors requiring different services with different priorities in space and time. To that aim, the network slicing based on virtualisation techniques makes it possible to modify the network behaviour by changing functions or reconfiguring parameters.

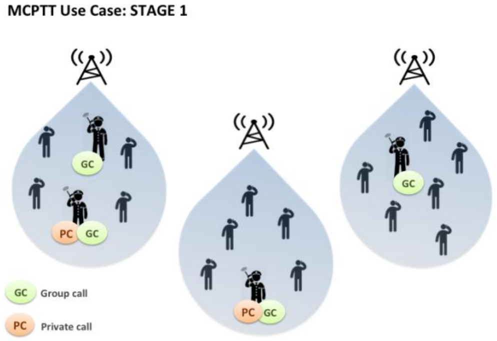

Figure 3: Stage 1 – Default service agreement.

In the first stage, we assume a normal situation with no detected incident. In normal circumstances, only a few connections are necessary to support a group call or a private call. Consequently, a Public Safety (PS) tenant only needs a reduced E2E slice to operate appropriately, and this slice shares network resources with other PS slices or commercial slices. Figure 3 shows an example of this situation for the MCPTT service.

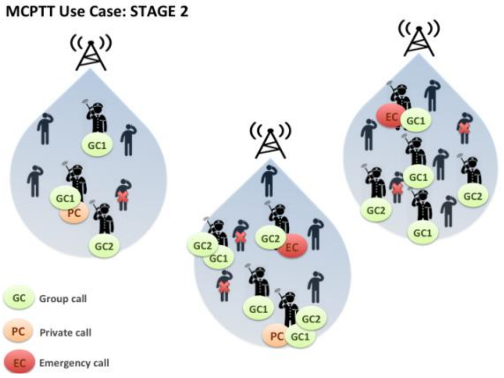

Figure 4: Stage 2 - Incident.

When an emergency incident occurs, it attracts more first responder personnel to the affected area. This fact involves the request of more or bigger call groups, more private calls and, in contrast to the stage 1, also emergency calls. Figure 4 depicts this situation for the MCPTT service. As previously commented, prioritised call requests must always be served. In consequence, the capacity of the PS slice should be increased. To that aim, the monitoring/telemetry system must be able to detect this situation and inform the NMS in order to reconfigure the resource allocation for the E2E slice through the NFV orchestrator, the EMS and, eventually, the cSD-RAN controller.

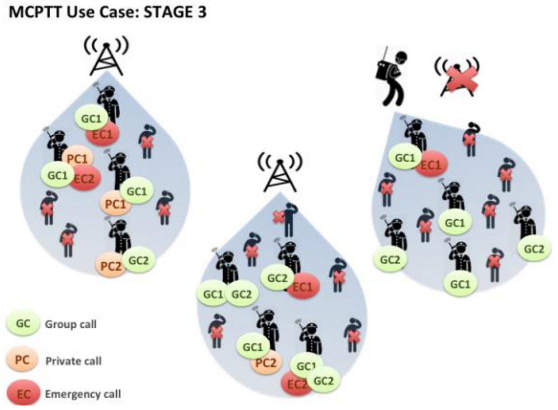

Figure 5: Stage 3 – Damaged infrastructure.

Finally, if the ICT infrastructure of the CESC cluster is damaged in the incident, even affecting the backhaul toward the core network, the stage 3 considers the addition of a deployable system for coverage extension. Figure 5 shows an example of this situation for the MCPTT service. Each deployable unit should be included in the cluster as a new CESC. In such a situation, where the shared resources are scarce and unstable, the available network infrastructure will be devoted exclusively to provide connectivity to first responders. In consequence, commercial network slices disappear, and MC services must be able to operate in isolation mode as tactical bubbles.

In the context on 5G ESSENCE these decisions will be made with the support of the telemetry & analytics module based on the collected monitoring and telemetry data. The isolation between different network slices is also a crucial aspect to consider, in order to guarantee communication security premises of mission critical services. In addition, both the software-defined mobile networks and the NFVI edge/core capabilities will be leveraged.

V. Cloud Resource Allocation

When building the cloud, resources need to be allocated and managed. The orchestrator plays an important part in that role. In the context of the 5G ESSENCE project, the proposed orchestrator is the ETSI Open Source MANO (OSM) [5]. The resources that need to be allocated and managed for the composition of the service, are both computing and networking resources.

The orchestrator, will delegate the task of managing the computing resources to the Virtual Infrastructure Manager (VIM), Open Stack [6] has been selected as emerging and reliable software for NFV infrastructure. Open Stack will allow the possibility of assigning slices for different tenants (guaranteeing tenant isolation), and also manage parameters associated to those slices, in particular to prioritize the bandwidth per slice.

Open Stack will also manage the service placement, and for that the concepts of compute nodes and availability zones are very important. In a nutshell, Open Stack will create virtual resources in a physical server, creating in that way a compute node where to deploy the services. If needed, these compute nodes can be grouped into availability zones, abstracting in that way the placement work, and delegating the placement decision to Open Stack. This concept is one of the key enablers for the two-tier orchestration. In same line, the concept of availability zones will be key in the case of 5G ESSENCE and the described scenario. Open Stack will be taking care of managing the resources added during the “Dynamic Addition of Deployable CESC”.

Finally, when talking about the networking resources, and in the context of 5G ESSENCE the orchestrator will have two options: either to rely on Open Stack for doing the basic service related networking, or then directly interacting with the SDN controller to perform more advanced tasks for managing data flows.

VI. Radio Resource allocation

The network slicing capability of 5G systems allows configuring different logical networks on top of the same 5G network to provide specific network characteristics and capabilities, offering a differentiated network behaviour to the User Equipments (UEs) of each slice. Therefore, as it has been previously discussed, Mission Critical (MC) services in 5G ESSENCE will be typically supported through a specific network slice, different from the one used for supporting other commercial services (e.g. eMBB). At the RAN, the support of different RAN slices involves different Radio Resource Management (RRM) functions to ensure that each RAN slice gets the expected amount of resources.

To illustrate the operation of these RRM functions, let us consider a scenario with two slices, one for a commercial operator (tenant 1) and the other one to a Public Safety operator (tenant 2). Both slices provide only Guaranteed Bit Rate (GBR) services with the characteristics given in Table I. They are defined in terms of the GBR value and the Allocation and Retention Priority (ARP) [7], which is a parameter that defines the relative importance of a service and allows prioritizing the different services in case of resource limitations. The scenario assumes one CESC with one channel of 50 Physical Resource Blocks (PRB) of 360 kHz, corresponding to one of the 5G New Radio (NR) numerologies [8]. Assuming a spectral efficiency of 5.6 b/s/Hz, the achievable data rate in one PRB is 2 Mb/s.

In this scenario with only GBR services, the main RRM function impacting on the RAN slicing is the Radio Admission Control (RAC). The RAC is executed whenever a new GBR service has to be established for a UE to check if it can be admitted or not. For that purpose, the RAC considers the current PRB occupation in the CESC, and GBR and ARP values of the different services. Besides, for properly isolating the two slices, the RAC condition considers a limit in the maximum allowed PRB occupation for all the admitted service requests of each slice.

Table I Services per Slice

Slice/Tenant	Service	ARP	GBR
1.- Commercial operator	Premium – Video HD	2	10Mb/s
	Basic - Video	3	3Mb/s
2.- Public safety operator	MC Video	2	5Mb/s
	MCPTT	1	48kb/s

The performance is evaluated under two different operational conditions in line with the stages 1 and 2 discussed in Section IV. Figure 6 depicts the blocking probability (i.e. the probability that the RAC rejects a service request) experienced by each service under normal conditions in which the traffic coming from the public safety operator is low while the traffic from the commercial operator is progressively increased. In this case, the maximum PRB occupation considered by the RAC is 60% for slice 1 and 20% for slice 2. In turn, Figure 7 depicts the blocking probability under a critical situation in which an emergency incident has occurred and, therefore, the public safety traffic is progressively increased whereas the traffic from the commercial operator remains moderate. To cope with this situation, the RAC is configured with a maximum PRB occupation of 30% for slice 1 and 50% for slice 2.

For the case of normal conditions, it can be observed in Figure 6 that the blocking probabilities of eMBB services grow when the offered load of tenant 1 is increased. In turn, the blocking probability of both MC Video and MCPTT services remains low and independent from the offered load of tenant 1, which reveals the isolation between the two slices. Instead, under the emergency conditions of Figure 7 it is observed that, when increasing the load of tenant 2 the MCPTT blocking probability remains at very low values while the MC Video blocking probability increases. This is due to the lower ARP (i.e. higher priority) associated to the MCPTT service. Again, proper isolation between both tenants is observed, i.e. the performance of tenant 1 services does not change when increasing the load of tenant 2.

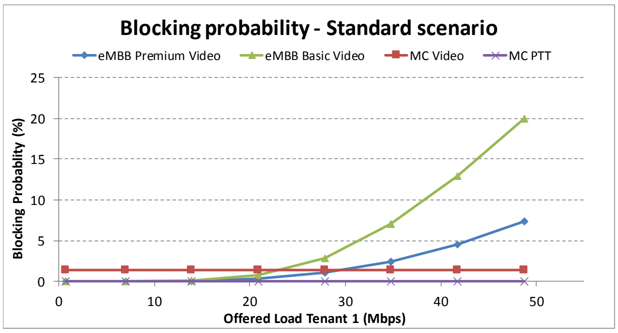

Figure 6: Blocking probability under normal conditions

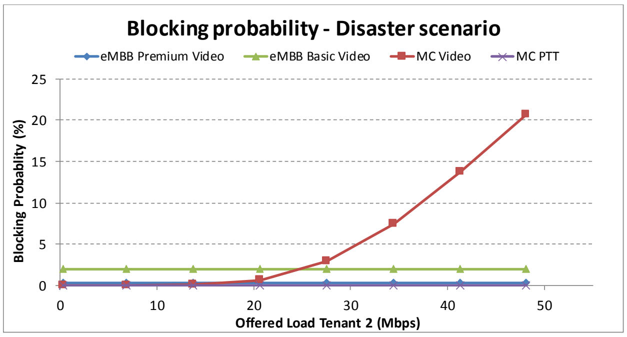

Figure 7: Blocking probability under emergency conditions

VII. QOS ASSURANCE

From a quality of service (QoS) perspective Mission Critical applications expect very reliable communications, very high availability and special operational needs. Those qualities are essential to make a mobile network compliant with critical communication systems. Due to this, MC Communication systems demand services and tools with specific requirements and functionalities, where communication failures are on high risk. Monitoring traffic and usage of the application is crucial to ensure the QoS, Radio and network resources need to be guaranteed, assigning higher priorities on the network provisioning than to consumers or enterprise users.

Mobile Edge Computing (MEC) also provides the capability to deploy services closer to the end user. To this end, Network traffic is reduced and communications capabilities such as security, low-latency and analytics application are improved. The user location information is important to deploy or migrate services closer to the user. the Resource Landscaper keeps track of physical and virtual resources available in a DC, including hardware/software ingredients and components, and combines this with contextual information that includes specific features of the resources and expresses their relationship with the services deployed on the infrastructure over time, to make more informed decisions on resource allocation.

The service environment context can provide valuable information that can be used to adapt and reconfigure the assigned resources. On the same line, MNOs should guarantee that traffic prioritization works correctly for private consumers and mission critical users.

The monitoring system used to track the infrastructure iis Prometheus [9], which allows general and tailored probes to monitor the infrastructure (servers, switches, physical radio parameters) and the behaviour of the network. It also provides an alerting system that allows notifying management systems of the infrastructure about potential problems in the platform. The system is scalable enough to automatically adapt when a new component is included in the infrastructure such as the inclusion of new Light DCs in the system, or if some of the components are not available or already in use.

The system is able to identify an emergency situation, i.e. number of connected users is increasing, and consequently proactively redistribute the assignment of the resources so the network does not get congested. Moreover, it will need to take the decision to limit or block the communication with concurrent services in the parallel communication slices to assure latency, reliability, isolation and connectivity on a 24/7 basis to prevent any degradation of these critical QoS parameters.

Taking all these considerations, using key quality assurance techniques such as proactive closed-loop automation, user tagging service assurance and service orchestration as well as automated problem identification analysis among others, MNO can begin to offer the challenging QoS values that MCS requires.

VIII. CONCLUSIONS

This paper has presented the management of network services to provide MCS for public safety to building on the current developments made by the 5G ESSENCE project. The cloud, network and radio resources need to be assigned to offer reliable capabilities in such a challenging environment. The work now will be focusing on enhancing the resource placement and radio configuration based on monitoring information to guarantee optimal services and applications in mission critical situations and for public safety.

IX. ACKNOWLEDGMENTS

The research leading to these results has been supported by the EU funded H2020 5G-PPP project 5G ESSENCE under the grant agreement no 761592.

X. REFERENCES

[1] https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx

[2] http://www.5g-essence-h2020.eu/

[3] https://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing

[4] 3GPP TS 22.179 v16.1.0, Mission Critical Push to Talk (MCPTT) over LTE; Stage 1 (Release 14), April, 2018.

[5] OSM Release FOUR Technical Overview

[6] OpenStack: The Path to Cloud

[7] 3GPP TS 23.501 v15.2.0 "System Architecture for the 5G System; Stage 2 (Release 15)", June, 2018.

[8] 3GPP TS 38.300 v15.2.0, “NR and NG-RAN Overall Description; Stage 2 (Release 15)”, June, 2018.

[9] Prometheus - Monitoring system & time series database