5G System Support for Mission Critical Communications

Evelina Pencheva[1], Aleksander Nametkov[1], Denitsa Velkova[1], Ventsislav Trifonov[1]

Abstract – **Fifth generation (5G) mobile networks will provide flexibility to customize quality of service for diverse use cases. This enables a wide range of mission critical services, including autonomous vehicles, industrial control, robotics, telesurgery, augmented/virtual reality, etc. These services require very high reliability and availability, and low latency. In this paper, we present the 5G developments to support mission critical communications. Topics discussed in brief include advances in radio access, network slicing, network programmability, and Device-to-Device communications. Utilization of Multi-access Edge Computing in 5G is considered in the context of mission critical services.**

Keywords – **Latency, Reliability, Radio access, Network slicing, Quality of Service, Multi-access Edge Computing.**

I. Introduction

As to 3GPP definition, mission critical communications require high reliability and availability, low setup and transfer latency, ability to handle large number of devices, priority handling and strong security. The mission critical communications serve public safety, utility sectors, railways and other vertical segments [1]. Historically, this type of communications have been provided by dedicated private mobile radio technologies (e.g. TETRA, GSM-R, iDEN, etc.) which are low capacity, narrowband, fragmented and are de – facto standards. The demand for broadband data applications and location-based services, and the aim to achieve high efficiency make four generation Long Term Evolution (LTE) and LTE-Advanced (LTE-A) reference technology for mission critical applications. LTE is widely supported by mobile network operators which may provide mission critical service by their commercial LTE networks [2], [3]. In comparison with private mobile radio technology, LTE provides higher data rates and better security mechanisms. However, LTE is not designed to comply with reliability, confidentiality and security requirements of mission critical services.

The upcoming 5G systems promise to support the growing needs for enhanced Broadband Communications (eBBC), massive numbers of connected devices (mMTC), and ultra-reliable, low latency communications (URLLC). URLLC services expose strict requirements on low latency and reliability for mission critical communications. Ultra-reliability and low latency are vital for applications such as intelligent transportation, telesurgery, industry automation etc. In this paper, we outline the vision and requirements of mission critical communications in 5G, and discuss 5G technologies that aim to support mission critical applications.

[1]The authors are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: enp@tu-sofia.bg.

The rest of the paper is organized as follows. Next section describes the different types of mission critical services as defined by 3GPP and their requirements. Section III discusses 5G developments aimed to achieve ultra-reliability and low latency. Section IV describes the 5G quality of service mechanisms for provisioning of high availability of mission critical services. Section V presents the capabilities of Multi-access Edge Computing to support mission critical applications. The conclusion summarizes the contribution.

II. Mission Critical Service Types and Requirements

3GPP defines three types of mission critical services [4]. The Mission Critical Push-to-Talk (MCPTT) service supports communication between several users (i.e. group call), where each user has the ability to gain access to the permission to talk in an arbitrated manner. The MCPTT service also supports private calls between two users. The Mission Critical Video (MCVideo) services and Mission Critical Data (MCData) can be used for public safety applications and maritime safety applications and also for general commercial applications (e.g. utility companies, railways and maritime usage). MCVideo defines a service for mission critical video communication using 3GPP transport networks. MCData defines a service for mission critical data services which needs to provide a means to manage all data connections of mission critical users in the field and provide relevant resources to the ones who need it.

Mission critical (MC) refers to meeting the needs of agencies providing Public Safety services such as, but not limited to, Police, Fire department, and Paramedic services. Those needs include high reachability, availability and reliability of the service, low latency, real-time operating capabilities, highly secured operations, inter-operability with other services and systems, private and group communications, handling of emergencies and ability to provide prioritization, pre-emption, queuing and quality of service (QoS). Other examples of MC use cases include autonomous vehicles, industrial machineries in smart factories, virtual/augmented reality, remote early warning sensors, remote radar system and many others.

End-to-end (E2E) latency requirements need to be less than 1 ms and can be achieved by faster, more flexible frame structure and new non-orthogonal uplink access. Ultra-high reliable transmissions can be time multiplexed with nominal traffic through puncturing, and additional base stations for public safety coverage. Ultra-high reliability requires packet error rate (PER) less than 10^{-9} and can be achieved by simultaneous links to both 5G and LTE for extreme mobility and fault tolerance. E2E security is provided by security extensions to air interface, core network and service layer. Fig.1 shows the requirements of some MC use cases.

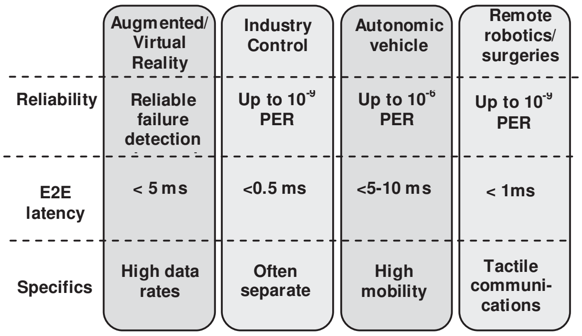

Fig.1 Requirements of mission critical use cases [5]

Common requirements to different types of MC services are defined in [5]. Efficient group communications are essential for different professional organizations. Geographic groups are based on the location and allow communication in a certain area. Functional groups are created for specific purposes and have to be available anywhere in the network. Fit-to-purpose groups are dedicated to specific tasks within a certain area. Efficient group management is required.

A mission critical system must prioritize different communication groups, e.g. Emergency Group Communications and Imminent Peril Communications provide users with higher priority in resource allocation. Users may have different mission critical applications, e.g. MCPTT, MCVideo, and/or MCData. These applications have different requirements for traffic handling, so users can have different prioritization levels for different applications. Furthermore, network may experience high level of traffic in extreme conditions which are unpredictable, and it must be able to adapt to exceptional MC load.

Simultaneous registrations of multiple devices have to be supported and the service administrator has to be able to limit the number of concurrent logs. Provisioning of location information is essential for MC services. Security requires support the confidentiality and integrity of all user traffic and signalling at the application layer.

III. Developments for Mission Critical Communications in 5G

The 5G introduces new developments to address the requirements of MC communications.

A. Radio Access

The radio access developments for ultra reliability include redundancy by usage of Massive Multiple-Input, Multiple-Output (MIMO) and multi-connectivity. Massive MIMO technology groups the antennas at the transmitter and at the receiver to achieve better spectral efficiency and throughput. 5G Radio Access Network (RAN) will support multi-connectivity for tight interworking among different 5G radio variants (below and above 6GHz carrier frequency), between 4G and 5G, and between different transmission points and carriers of the same radio variant [5].

Low latency is achieved by service-specific optimization of the protocol stack. Multi-connectivity and carrier aggregation are important for ultra-high reliability. In 5G, the functionality of Packet Data Convergence Protocol (PDCP) and Radio Link Control (RLC) are grouped in upper layer, and Media Access Control (MAC) forms lower layer which results in processing and latency gain due to less header processing and function optimization. Multi-connectivity and carrier aggregation is important for URLLC. With the proposed layering upper layer is able to connect to multiple lower layer entities, where the re-transmissions may be done at PDCP level. Further, segmentation is not required due to small and typically fixed packet sizes. Ciphering may be skipped if application level security mechanisms are applied for mission critical service.

B. Network Slicing

Both the 5G RAN and core network should support network slicing. Network slicing is introduced to support the required flexibility in provisioning of QoS required by different services [6]. Network slices are logical networks deployed over the same physical infrastructure. Each network slice provides specific network capabilities and network characteristics. For example, URLLC specific RAN functions include optimal handling of Radio Resource Control states to reduce state change latency, applying acknowledge mode only, prioritization of Random Access Channel, optimized coding for short payloads, and potential omitting of ciphering and header compression.

Network slicing supports slice isolation which prevents from distribution fault or security related events in one slice into other slices.

C. Network Vitualization

Network virtualization means network transition from hardware mode to software mode.

Both 5G RAN and 5G core network designs adopt Software Defined Networking (SDN). SDN splits the control plane and data plane. SDN may be applied to the following types of network functions: network control functions (session management, mobility management and QoS control), connectivity management (packet forwarding) and wireless control functions (such as scheduling and radio link adaptation). These network functions are performed by programmable and logically centralized controller which provides network technology independence.

5G networks will adopt Network Function Virtualization (NFV) also. NFV is a concept for replacing dedicated network devices, such as routers and firewalls, with software running on general purpose servers. It optimizes creation, activation, and provision of services by using the cloud advancements.

D. Device-to-Device Communications

The increasing demand for multimedia communications with high bandwidth and low latency requirements becomes a challenge for cellular networks. A new paradigm that may face this challenge is Device-to-Device (D2D) communications. D2D communications allow devices in close proximity to communicate directly without involvement of network elements processing of traffic data. D2D communications may address ultra-reliable and low latency requirements through direct communications between devices and thus reducing latency [7]. For MC communications, it is essential to ensure that the communication service is also provided if the network or parts of it are congested. D2D communication may address also reliability. In case of network congestion, a device with good internet connectivity may cash the data and transmit it to other devices. The network coverage may also be expanded by a device acting as a relay between the base station and other devices.

IV. 5G QoS Mechanisms for High Service Availability in 5G

In order to support a large diversity of use case requirements, 5G has to exploit flexible and highly granular end-to-end means to prioritize different traffic types, and to handle appropriately the packets belonging to the same flows. As an example for the latter, it is necessary to assign higher priority to short packet transferring protocol requests and responses than the other protocol packets, in order to avoid delays in connection setup time.

With respect to this context the key limiting factor of LTE QoS architecture is that the finest granularity to distinguish between data is on radio bearer level. Fig.2 shows the QoS architecture of 4G. Evolved packets System (EPS) bearer represents a level of granularity for QoS control which provides a logical transmission path with well defined QoS properties. The EPS bearer is mapped to QoS concepts of the underlying transport. Each EPS bearer is transported over the radio bearer with corresponding QoS characteristics. The EPS bearer QoS is mapped onto Internet protocol (IP) transport layer QoS between Service gateway (SGW) (E-UTRAN radio Access Bearer - ERAB) and Packet Data Network Gateway (PGW) (S5/S8 bearer). The mapping is one-to-one, which means that one radio bearer has to carry data related to different service requirements, but can not treat the packets related to one session differently. Hence, a higher level of QoS granularity is required.

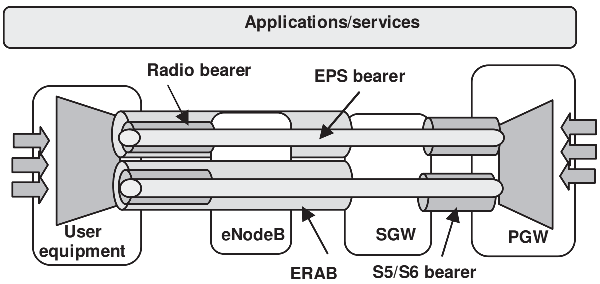

Fig.2 QoS architecture in 4G

The QoS architecture in 5G is based on the concept of QoS flow, which is identified by QoS flow ID (QFI) and is managed by Session Management Function (SMF) in the core network [8]. In contrast to 4G, in 5G, multiple QoS flows may be assigned to one PDU session. Each QoS flow is characterized by QoS profile, which is provided in a form of QoS rules by SMF to the User equipment (UE). SMF provides to the User Plane Function (UPF) in the core network QoS identifier (5QI) which contains information about guaranteed and non-guaranteed bit rates, and Allocation and Retention Priority in downlink and uplink, respectively. QoS flows can be flexibly assigned to different data radio bearers. The UPF applies service data flow classification in downlink, and the UE evaluates the QoS rules provided by the SMF and assigns the appropriate radio data bearers. Fig.3 shows the QoS architecture in 5G.

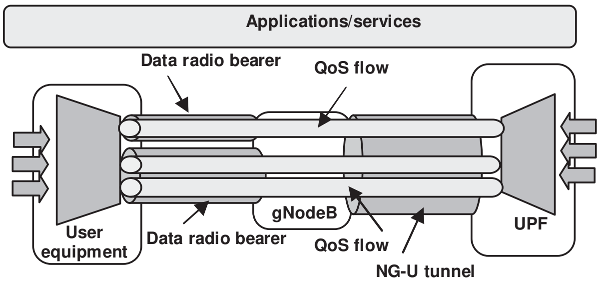

Fig.3 QoS architecture in 5G

V. Multi-access Edge Computing

The E2E latency may be reduced by deployment of Multi-access Edge Computing (MEC). MEC extends the cloud capabilities close to the place where they are used. It introduces data centers at the network edge. MEC can improve user's quality of experience (QoE) and can guarantee maximum utilization of RAN resources. The vicinity to end users enables applications with high bandwidth and low latency requirements. The MEC server contains the mobile edge platform and a virtualized infrastructure which provides cloud intelligence for the mobile edge applications. The mobile edge platform offers an environment for mobile edge applications to discover and to use available mobile edge services, and handles the user plane according traffic rules. Mobile edge applications run in a well-isolated manner on the virtualized infrastructure. Different MEC applications hosted by the MEC platform may belong to different network slices configured in RAN and/or core network.

MEC service platform offers two types of services. Radio network information services provide authorized applications with low level real-time radio and network information related to users and cells. Traffic offload function prioritizes traffic and routes the selected, policy-based, user data stream to and from applications that are authorized to receive the data.

There are many potential scenarios for MEC deployment [9], [10]. Edge may be referred to both RAN nodes and aggregation points in distributed core networks which serve specific requirements. For example, the MEC platform and applications may sit on an entity in the RAN for device offloading scenarios, while the MEC platform and applications may reside close to dedicated distributed core functional entities for scenarios where mobility and session continuity support is required.

The best way to receive radio network and location information and to manage the bandwidth is the "Bump in the wire" scenario where MEC is deployed much closer to RAN [10]. This scenario exposes the benefit from all MEC advantages such as low latency, efficient bandwidth management and local breakout. It is possible to bundle MEC platform and gNB (5G base station) into a single node where MEC can share the same network function virtualization infrastructure with Cloud Radio Access Network (C-RAN). When MEC is deployed in proximity of gNB or at aggregation point, the MEC platform is located on the N2 reference point between gNB and AMF, while the MEC data plane sits on the N3 reference point between RAN and UPF.

Fig.4 illustrates the MEC and C-RAN co-location deployment scenario. Fig.5 shows standalone MEC deployment at aggregation point.

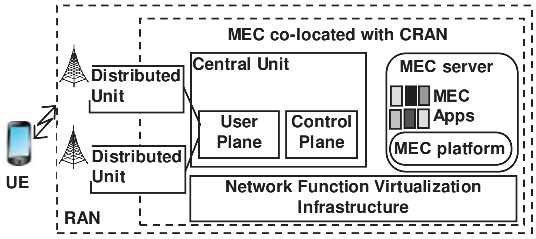

Fig.4 MEC and C-RAN deployment scenario

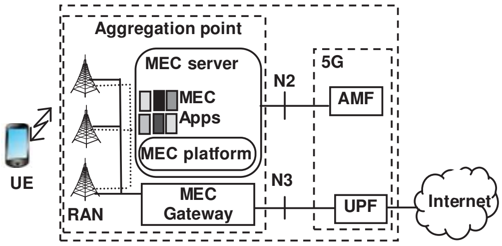

Fig.5 Standalone MEC deployment at aggregation point scenario

The MEC may be co-located with distributed core network functions (i.e. at the same site) [11], [12]. In this case, MEC applications may steer user plane traffic. This type of deployment is appropriate for MCPTT where communication with the core site is optional, as the traffic does not need to pass the backhaul to keep service running. This type of MEC deployment might usually be used by public safety, first responders, and mission critical industrial sites. Fig.6 shows the scenario.

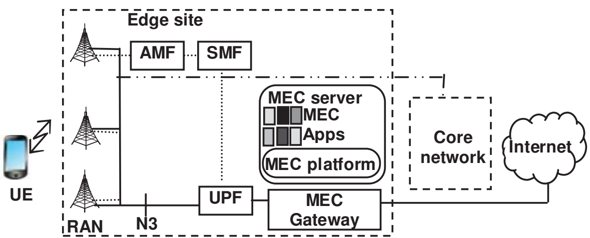

Fig.6 MEC deployment scenario with distributed core functionality

VI. Conclusion

In this paper, we consider challenges and developments for provisioning of mission critical communication in 5G networks. The requirements for ultra-reliability and low latency may be addressed by the advanced function exposed by 5G radio access, network slicing, network virtualization, D2D communications, enhanced QoS architecture and deployment of cloud computing at the network edge. 5G radio access supports multi-connectivity and protocol stack optimization. Network slicing, programmable networks and MEC further contribute to latency reduction and provisioning of full mobility and highest reliability.

Acknowledgement

The research is conducted under the grant of project DH07/10-2016, funded by Bulgarian National Science Fund, Ministry of Education and Science.

References

[1] Z. Kaleem, et al. "UAV-Empowered Disaster-Resilient Edge Architecture for Delay-Sensitive Communication", Cornell University, Networking and Internet Architecture, arXiv:1809.09617v2 [cs.NI], January, 2019.

[2] K. O. Olasupo, I. Kostanic and T. O. Olasupo, "Performance evaluation of mission critical communications services over LTE networks," IEEE ICPCSI, Chennai, 2017, pp. 273-278.

[3] J. Oueis, V. Conan, D. Lavaux, R. Stanica and F. Valois, "Overview of LTE Isolated E-UTRAN Operation for Public Safety," IEEE Communications Standards Magazine, vol. 1, no. 2, 2017, pp. 98-105.

[4] 3GPP TS 23.179 Functional architecture and information flows to support mission critical communication services, Stage 2, Release 13, v13.5.0, 2017.

[5] 3GPP TS 22.280 Mission Critical Common Requirements (MCCoRe), Stage 1, Release 16, 2018

[6] E., Pencheva, I. Atanasov, D. Kireva, K. Nikolova, "Network Slicing: A Mobility Management Perspective," SGEM 2018, vol.18, Green Design and Sustainable Architecture, pp.641-648

[7] A. Orsino et al., "Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical MTC in 5G," IEEE Communications Magazine, vol. 55, no. 2, pp. 79-87, 2017.

[8] Q. Ye, J. Li, K. Qu, W. Zhuang, X. S. Shen and X. Li, "End-to-End Quality of Service in 5G Networks: Examining the Effectiveness of a Network Slicing Framework," IEEE Vehicular Technology Magazine, vol. 13, no. 2, pp. 65-74, 2018.

[9] S. Kekki, et al. "MEC in 5G networks, " ETSI white paper no 28, June 2018.

[10] F. Guist, et al. "MEC Deploymnets in 4G and Evolution Towards 5G, "ETSI White paper, 2018.

[11] P. Skarin, W. Tärneberg, K. Årzen and M. Kihl, "Towards Mission-Critical Control at the Edge and Over 5G," 2018 IEEE International Conference on Edge Computing (EDGE), San Francisco, CA, 2018, pp. 50-57.

[12] R. Solozabal, A. Sanchoyerto, E. Atxutegi, B. Blanco, J. O. Fajardo and F. Liberal, "Exploitation of Mobile Edge Computing in 5G Distributed Mission-Critical Push-to-Talk Service Deployment," IEEE Access, vol. 6, 2018, pp. 37665-37675.