Public Safety Mobile Broadband System: From Shared Network to Logically Dedicated Approach Leveraging 5G Network Slicing

Anuar Othman and Nazrul Anuar Nayan

***Abstract*—The fourth generation (4G) of the mobile broadband system has widely been considered as the basic technology for the evolution of public safety communications system from legacy narrowband technologies. While 4G provides the essential features for the realization of a public safety network, it has limited flexibility and scalability due to its reliance on hardware-based network functions and its lack of intelligent and automated control capabilities. The fifth-generation (5G) standard introduces the sophisticated concept of network slicing as one of its design principles. The concept enables the creation of logical networks on a common infrastructure with appropriate isolation, resources, and optimized topology to serve a particular use case. In this article, studies on network slicing conceptual framework are summarized. Moreover, the characteristics of a public safety network alongside the current solutions in the 4G system to realize such characteristics are discussed. Furthermore, recent literature that proposes enhancements to 5G system functionalities, which are essential for the efficient and agile management of diverse resources, is examined. In conclusion, substantial research efforts are necessary to improve these functionalities to meet the rigid requirements of public safety use cases in terms of network slice reliability, resiliency, and security.**

***Index Terms*—Fifth-generation (5G) network slicing, fourth-generation (4G) system, machine learning, mission-critical services, network function virtualization (NFV), public safety network, software-defined networking (SDN).**

I. Introduction

PROFESSIONAL mobile radio (PMR) users from public safety agencies and industries have been relying on private networks and specialized technologies, such as terrestrial trunked radio (TETRA) and Project 25, for their communication services. These technologies have a proven track record in providing reliable, resilient, and secure networks required by demanding public safety users. However, such technologies are narrowband; hence, their capabilities for broadband services are limited [1]. A mobile broadband system provides several advantages, including the deployment of new services, such as video streaming and augmented reality, and migration of legacy services, such as group voice call and group text messaging. These new capabilities may greatly benefit public safety agencies in terms of the improvement of their operational efficiency and the safety of officers and citizens. The fourth generation (4G) of the mobile broadband system and its future evolutions have widely been envisaged as the basic technology for the evolution of public safety communication systems [2]. The current 4G technology is based on the long-term evolution-advanced (LTE-A) standard that has been defined by the third-generation partnership project (3GPP) since 3GPP Release 10. Similar to its predecessor, long-term evolution (LTE), the architecture of LTE-A, includes evolved packet core and evolved universal terrestrial radio access network (E-UTRAN) domains. Moreover, the standard offers improved data rate, spectral efficiency, user throughput, and link latency [3]. There are already several ongoing 4G-based public safety broadband projects, such as U.S.' FirstNet, U.K.'s ESN, and South Korea's SafeNet.

The deployment of a public safety broadband system can be undertaken in different approaches. The system can be deployed as a dedicated network owned by agencies, as a shared network with commercial operators, or as a combination of both [4]. Most ongoing projects adopt the combined approach rather than building a fully dedicated one because of the cost efficiency and faster time to market of the former. The dedicated network approach provides full control of the network, subscribers, and communications to public safety agencies. However, this approach requires a dedicated spectrum, which poses challenges because, in certain countries, the spectrum allocation is managed using market-based mechanisms to provide an optimal contribution to the national economy [5]. All these approaches must satisfy rigid requirements in terms of network reliability, resiliency, and security. Moreover, public safety users normally work in groups and, therefore, require a system that provides efficient support for group-based communications services [6]. In addition, these services must always be accessible in all scenarios, including during emergencies and disasters when network infrastructure is frequently degraded or destroyed. To this end, 4G is continuously enhanced with public safety-related features, such as the ones that enable the deployment of group communications and allow direct communication between user devices. Furthermore, in a shared network approach, 4G permits the prioritization of public safety services over the ones used by

Manuscript received December 18, 2019; revised April 13, 2020; accepted June 5, 2020. This work was supported by the Universiti Kebangsaan Malaysia Research under Grant GUP-2019-020. *(Corresponding author: Nazrul Anuar Nayan.)*

The authors are with the Department of Electrical Electronics and Systems, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia. (e-mail: p97887@siswa.ukm.edu.my; nazrul@ukm.edu.my).

Digital Object Identifier 10.1109/JSYST.2020.3002247

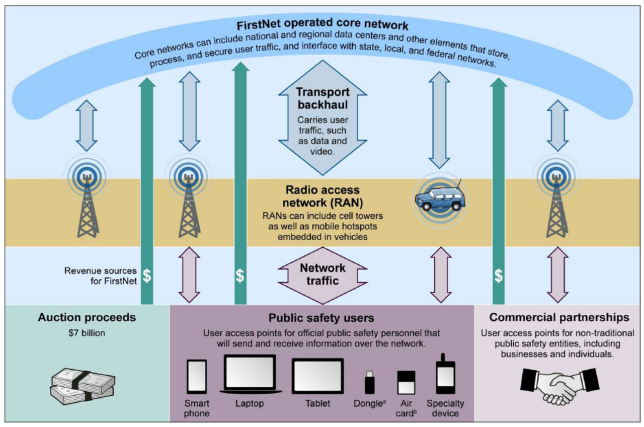

Fig. 1. Key elements of the FirstNet public safety network [7].

the consumers. FirstNet is one of the public safety broadband systems that combine this approach with a dedicated network [7]. As illustrated in Fig. 1, FirstNet consists of core networks, radio access network (RAN), and transport backhaul. The fixed base stations of dedicated RAN are deployed in areas that require greater capacity, while the deployable ones are exploited in underserved areas or during disasters. The other areas are served by existing RAN that is shared through commercial partnerships [8]. In brief, 4G provides the essential features for the successful deployment of public safety broadband systems.

The advent of the fifth-generation (5G) mobile broadband standard introduces additional sophisticated technologies that may further benefit public safety users. One of them is network slicing that enables the creation of logical networks on a common infrastructure with appropriate isolation, resources, and optimized topology to serve a particular use case [9]. The 3GPP Release 15 provides the first full set of 5G standards, which include support for network slicing. Nakao *et al.* [10] provide a summary of the preliminary research effort on network slicing and discuss use cases in different categories, namely, massive machine-type communications, enhanced mobile broadband, and ultrareliable and low-latency communications. Ordonez-Lucena *et al.* [11], [12] focus on the architectural aspects of the 5G system to enable an efficient implementation of network slicing and provide the realization options and deployment examples. In a previous article [13], we review resource allocation schemes for 5G network slicing from the perspective of public safety use. These schemes, leveraging artificial intelligence techniques, are proposed in different network domains and at the various stages of the network slice lifecycle. In summary, network slicing may profit public safety network deployment in terms of cost reduction and faster time to market in comparison with a fully dedicated network. Furthermore, network slicing on a commercial network may provide greater control to network tenants, such as public safety agencies, than the 4G network sharing approach.

This article aims to investigate the network slicing conceptual framework and 5G functionalities that provide efficient support for its realization and customization for public safety use cases. The remainder of this article is structured as follows. Section II presents a review of the state-of-the-art architectural framework of the 5G system with a focus on network slicing support. Section III provides an overview of the public safety network characteristics and existing solutions in the 4G system to realize such characteristics. Section IV discusses the recent studies that propose enhancements to 5G functionalities, which are essential for the efficient and agile management of diverse resources to meet the specific requirements of a network slice. Finally, Section V provides the concluding remarks.

II. 5G Network Slicing

In this section, we briefly review the network slicing concept and its enabling technologies. We also review the 5G conceptual architecture proposed by industries and standards developing organizations (SDOs) and recent studies, which enhance this baseline architecture to facilitate its realization.

A. Concept and Use Cases

Future 5G system is expected to support different business models and use cases in addition to enhancing the current 4G performance. A study on future business requirements is conducted by the next generation mobile network (NGMN) alliance from the perspective of network operators. It emphasizes the need for a flexible mobile system with modular network functions to fully optimize resource usage while providing scalable and cost-efficient solution [14]. To this end, the NGMN alliance introduces the concept of network slicing as one of the 5G design principles. In [15], NGMN further describes the concept and provides its high-level architecture. An instance of network slicing is defined as a combination of network functions and their related resources, forming a complete logical network that can meet certain characteristics requested by a communication service. This instance, referred to in this article as a network slice, is created from a network slice template. Moreover, a network function refers to processing function in a mobile system, such as a packet data gateway and firewalls. Related resources of a network function can be categorized as physical and logical ones. The former consists of the physical asset for computation, storage, transport, and radio access, whereas the latter is composed of the partition or aggregation of these physical assets. Briefly, the network slicing concept envisioned by NGMN enables the 5G system to efficiently support different use cases with diverse characteristics on a common infrastructure.

Network slicing allows mobile operators to provide customized logical networks to third-party customers or tenants that can be business vertical or virtual operators with their own subscribers. Critical communications for public safety with rigid requirements in terms of network reliability and resiliency in all conditions are one of the use cases described by NGMN [14]. Moreover, Zhou *et al.* [16] introduce the network slicing as a service (NSaaS) business concept that enables the operators to provide cloud-based logical networks to its tenants. In addition, 3GPP recommends operators to provide network slice as an operator internal, which facilitates the on-demand provisioning of slices when requested by the customers [9]. For example, a public safety agency procures a logical network by using

the NSaaS concept from an operator and provides group voice communication service to its users. During special events, such as the general election or riot, the agency requests the operator to provide video streaming services by using another slice owned by the operator. In brief, network slicing allows the operators to offer innovative services to business verticals, thus providing them with new business opportunities.

B. Management and Orchestration

The recommendations for management and orchestration functions at the network slice instance layer are provided by 3GPP in [9]. A network slice lifecycle consists of several stages that starts with a preparation phase where its template is created. The template describes all functionalities and resources required to provide a communication service. Then, in the instantiation, configuration, and activation phase, all network functions and their related resources are created and configured before the network slice can be activated. In the next phase known as runtime, the network slice handles its services traffic while being supervised continuously. If needed, the network slice can be reconfigured and scaled accordingly. Finally, the network slice is deactivated, and its network functions and their related resources are reclaimed in the decommissioning phase. In addition, for management purposes, 3GPP introduces the concept of a slice subnet, which forms a part of the complete logical network. Subsequently, the network slice management involves communication service management function (CSMF), network slice management function (NSMF), and network slice subnet management function (NSSMF). Upon receiving a communication service request, CSMF first translates the request to network-slice-related requirements. Then, NSMF derives the slice-subnet-related requirements and passes the information to NSSMF. NSMF and NSSMF are responsible for the management and orchestration of network slice and slice subnet, respectively.

C. Enabling Technologies

Legacy mobile systems extensively rely on hardware-based network functions that run on the specialized hardware. These systems provide network services, such as a single-user connectivity, by statically connecting or chaining multiple network functions together [17]. To support network slicing, the 5G system must offer flexibility in terms of network function chain composition and placement, and allocation of related resources. To this end, the European Telecommunications Standards Institute (ETSI) created an industry specification group for network function virtualization (NFV). The principal objective of NFV is to enable the implementation of network functions as software modules running on commodity hardware [18]. Thus, the main benefits of NFV include capital expenditure reduction, operational efficiency improvement, and enabling of the multitenant network [17]. Furthermore, virtualization technology allows network function software to be virtualized and dynamically chained to provide a network service. This chaining is described by the virtual network function (VNF) forwarding graph. In summary, NFV enables a flexible 5G system, which is essential to realize customized network slices on a common infrastructure.

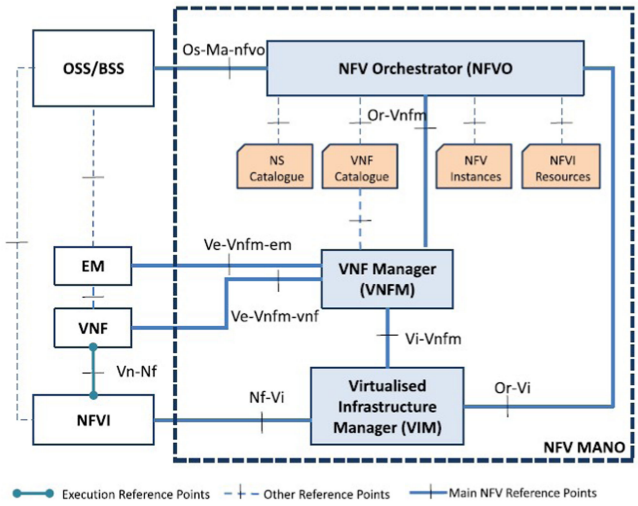

Fig. 2. NFV MANO architectural framework with reference points [20].

The high-level architectural framework of NFV is described by the ETSI in [19] and [20]. It contains three working domains: NFV infrastructure, VNFs, and NFV management and orchestration (NFV MANO). The physical and virtual resources for computation, storage, network, and virtualization layer are included in the NFV infrastructure, whereas NFV MANO handles the lifecycle management of VNFs and the network services that they form. As illustrated in Fig. 2, NFV MANO is composed of an NFV orchestrator, a VNF manager, and a virtualized infrastructure manager (VIM). VIM is responsible for the management of NFV infrastructure, whereas a VNF manager performs the lifecycle management task of a single VNF instance. The NFV orchestrator coordinates resources across multiple VIMs and handles the lifecycle management of network services. Element management system and operations and business support systems are the legacy management systems of operators, which are responsible to, respectively, manage VNF functionality and to perform fault, performance, configuration, and accounting management tasks. Briefly, NFV architecture provides comprehensive management for network service, its constituent VNFs, and related resources.

Network service programmability in the 5G system is also essential to enable agile and scalable network slice against varying service demands and network conditions in relatively short timescales. Legacy mobile systems have limitations in optimizing their resource allocation in dynamic environments due to the lack of intelligent and automated control capabilities. To this end, the open networking foundation (ONF) specifies software-defined networking (SDN) architecture in [21] with the objective of providing an open interface that allows a tenant or a client application to have exclusive control over a group of resources. In essence, SDN decouples the control planes of resources from its data planes and logically centralizes them. Relevant information exposed by data plane resources is then provided to client applications. The high-level SDN architecture

is composed of data, controller, and application planes. The controller acts as the intelligent node in a control loop to optimize resource allocation according to a predefined policy. It also translates requirements from the application to the dynamic and granular control over network elements. A network element represents a group of abstract data plane resources managed as a single entity. In summary, an SDN enables network service programmability through the decoupling and centralization of control planes and through resource abstraction and exposure to application.

The realization of network slicing may greatly benefit from 5G system flexibility and adaptability driven by SDN and NFV technologies. In [22], ONF argues that SDN resource virtualization can enable the logical isolation and sharing of 5G heterogeneous resources among different slices. In addition, SDN control plane centralization allows exclusive control over these resources by software applications that can be, for example, OSS at the tenant domain or a VIM at the infrastructure domain. Moreover, the open and programmable interface of the SDN controller facilitates dynamic resource allocation and optimization during the network slice lifecycle. ETSI explores the NFV architectural framework mapping into the network slicing concept in [23]. From the resource management perspective, a network slice with at least one VNF can be represented by a network service instance or a group of them chained together. Therefore, the management functions for network slicing and NFV interacting with each other, and the lifecycles of their instances are interrelated. The creation of a new network slice template can trigger the update of the existing network service descriptor or generation of a new one, while instantiation, configuration, activation, and decommission of a network slice can cause the same actions on its constituent network services. Furthermore, during the runtime phase, NFV MANO supervises the virtualized resources of a network slice and, if needed, reconfigure and scale them accordingly. In brief, SDN enables the efficient use of common infrastructure by multiple network slices that are dynamically optimized according to service requirements. Meanwhile, NFV facilitates the management of virtualized resources from different domains.

D. 5G Architectural Enhancements

Recently, network slicing has been eliciting increasing interest from academia, industries, and research institutes. Among the prominent research groups are the European union-funded 5G infrastructure public–private partnership (5GPPP) projects that provide consolidated output and view on the overall architecture to SDOs, such as 3GPP. Phase one of the 5GPPP projects focuses on designing the 5G baseline architecture at a conceptual level, which includes the support of network slicing. Phase two projects address the practical realization of this design.

With network slicing, the 5G system is expected to support diverse service portfolios and multiple service providers on the same infrastructure. The requirements for these multiservice and multitenancy support entail the application of SDN and NFV in the design of 5G system architecture. One of the 5GPPP phase one projects, 5G novel radio multiservice adaptive network

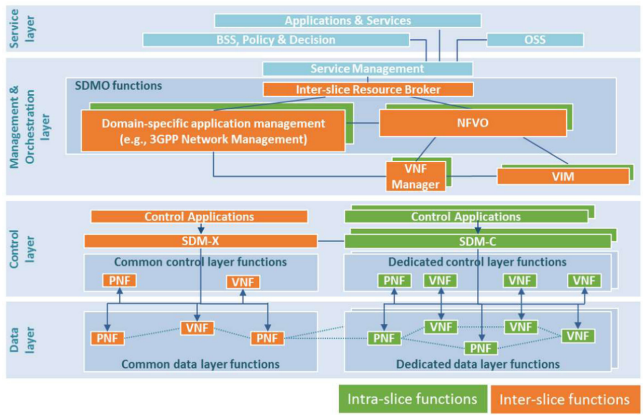

Fig. 3. Functional perspective of the overall 5G NORMA architecture [26].

architecture (5G-NORMA), extends NGMN architectural vision by leveraging these two technologies to enhance the 5G programmability and flexibility. Within the project scope, Rost *et al.* [24] highlight the importance of programmable architecture in enabling the multiservice networks and describe the SDN-based software-defined mobile network control concept that centralizes the control planes of network functions. In accordance with the grouping of network functions into the shared and dedicated slice subnets in [9], 5G-NORMA introduces software-defined mobile network controller (SDM-C) and software-defined mobile network coordinator (SDM-X) components to efficiently manage the network functions of the dedicated and shared slice subnets, respectively [12]. Moreover, Samdanis *et al.* [25] introduce the network slice broker concept, which is a centralized control component that replaces the passive sharing approach between the different tenants with on-demand resource allocation as well as the automated admission control and radio resource scheduling. As illustrated in Fig. 3, the proposed 5G-NORMA conceptual architecture employs VNFs and physical network functions, which are managed by NFV MANO and 3GPP network management, respectively [26]. A single interslice broker and SDM-X are deployed in a 5G network to, respectively, allow the optimized resource orchestration and dynamic sharing of common network function among different slices, thus enabling a multitenant system. Moreover, each network slice is assigned with an NFV orchestrator and an SDM-C, both of which are responsible for providing efficient management of network functions and resources dedicated to that slice, thus enabling multiservice support within a slice. Briefly, the 5G-NORMA conceptual architecture enables the flexible and agile 5G systems, which are capable of efficiently supporting multitenant and multiservice environments on the same infrastructure.

The realization of the network slicing concept requires further enhancements to the 5G system architecture, especially in supporting the concept's end-to-end framework, as envisioned by NGMN. One of the 5GPPP phase two projects, 5G-MONARCH, aims to bring into practice the conceptual architectures proposed by 5G-NORMA. Within the project scope, Gutierrez-Estevez *et al.* [27] identify several challenges in the baseline architecture, such as the lack of support for the graceful scaling of

VNF resources against dynamic traffic load and the need for interslice resource optimization across all domains and tenants. Another limitation is caused by the use of legacy protocol stack, especially in the RAN domain that requires certain network functions to be hosted on the same node for optimality reason. To address these key challenges, the authors introduce the concept of resource elasticity that is composed of three innovation areas: computational elasticity, orchestration-driven elasticity, and slice-aware elasticity. The first innovation area involves the design of automated VNF scaling mechanisms, while the second one focuses on the flexible mapping of a VNF chain over heterogeneous cloud resources. The third innovation concentrates on optimizing resource allocation among slices by leveraging adaptive multiplexing schemes. These innovations aim to improve resource usage efficiency at VNF, and intraslice and interslice levels, respectively. Moreover, Shariat *et al.* [28] propose three enabling innovations, namely, a cloud-enabled protocol stack, interslice control and management, and experiment-driven optimization. The first innovation relaxes the temporal and logical interdependencies among network functions through protocol stack redesign, thus enabling their flexible placement in different nodes. The second innovation enhances slice awareness in the RAN and core network domains. The innovation in the RAN domain consists of designing a two-level radio resource management functionality in which one optimizes resource allocation among slices in coarse timescales, whereas the other performs the same task within a slice at a finer granularity. The innovation in the core network domain involves the design of a core network using a service-based architecture approach, as defined in [29], in which the control plane network functions are interconnected via a common bus, thus facilitating interslice optimization. Finally, the third innovation refines the behavior models of VNFs by using measurements from the live networks, which can be used to improve resource orchestration algorithms. In brief, 5G-MONARCH provides necessary improvements toward bringing 5G conceptual architecture into practice.

As envisioned by NGMN, the 5G system must support end-to-end network slicing that includes devices, network infrastructure from different domains, and management functions. To this end, the architecture designed by 5G-MONARCH extends the ones proposed by 5G-NORMA, ETSI, and 3GPP with a new layered approach, internal functions, and interconnecting interfaces [30]. Similar to 5G-NORMA, in the controller layer, the dynamic controls of the dedicated and shared network functions are handled by an intraslice controller and an interslice controller, respectively. These SDN-based controllers interact with underlying network functions in the network layer via a southbound interface and provide their exposed information to the applications via a northbound interface. Moreover, in the management and orchestration layer, the inclusion of all technology and network domains in its lower sublayer enables full support for end-to-end network slicing. In its upper sublayer, the integration of 3GPP management functions, namely, CSMF and NSMF, facilitates the process of translating requirements from the service layer into slice requirements and subsequently enables the distribution of these requirements among different domains. The management and orchestration layer interfaces

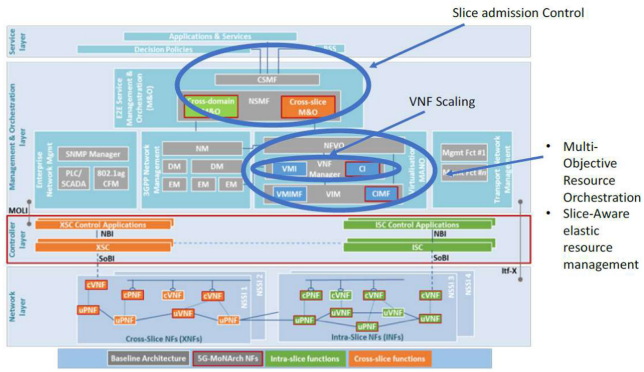

Fig. 4. Innovations on elastic network lifecycle management on the top of 5G-MONARCH architecture [31].

with the controller and network layers via specific reference points. These architectural enhancements further facilitate the implementation of the aforementioned resource elasticity concept. As illustrated in Fig. 4, 5G-MONARCH integrates the functional innovations at the management and orchestration layer to enable the elastic network lifecycle management [31]. When a network slice request is received from a tenant, the slice admission control function at the upper sublayer uses a policy to accept or reject the request based on the current system states. Then, the multiobjective resource orchestration function at the NFV MANO component optimizes the allocation of heterogeneous resources among admitted slices during the slice instantiation and configuration phases, whereas the slice-aware elastic resource management function performs the allocation of radio and computational resources to each slice. Next, throughout the runtime phase of each activated slice, the VNF scaling function at VNF manager scales or migrates the slice constituent VNF chains to meet the service requirements. In summary, the proposed architectural extension by 5G-MONARCH enables the support for end-to-end network slicing and the implementation of the resource elasticity concept.

III. Public Safety Broadband Network Characteristics

In this section, we highlight the characteristics of a public safety network and discuss existing 4G features to meet such requirements. We also review recent studies that propose enhancements to these features.

A. Group Communication Service

Public safety users normally work in groups and, therefore, require a system that provides efficient support for group-based communications. To this end, 3GPP introduces group communications system enabler (GCSE) in Release 12 [32] and enhanced multimedia broadcast multimedia multicast (eMBMS) in Release 10 [33]. The GCSE is a logical interface that facilitates the creation, management, and deletion of group communications, either through a centralized approach by a core network and an application server or a distributed approach by a base station.

TABLE I
KEY PERFORMANCE INDICATORS (KPIS) OF MISSION-CRITICAL SERVICES CONSISTING OF PUSH-TO-TALK VOICE (MCPTT) SERVICE [34]

[ ]

Variable	Meaning	Value
KPI 1	MCPTT Access time	< 300 ms
KPI 2	End-to-end MCPTT Access time	< 1000 ms
KPI 3	Mouth-to-ear latency	< 300 ms
KPI 4[a]	Maximum Late call entry time (without application layer encryption)	< 150 ms
KPI 4[b]	Maximum Late call entry time (with application layer encryption)	< 350 ms

The eMBMS is a multicast transmission framework that allows a single or more base stations to transmit the same downlink radio signal to multiple users. Moreover, eMBMS provides two transmission schemes, namely, single-cell and single-frequency networks. In the former, the base stations transmit multicast signals independently from each other. In the latter, a group of synchronized base stations transmits the signal simultaneously, thus improving spectral efficiency and radio coverage. Briefly, 4G provides essential features to efficiently support group communications for public safety users.

Legacy PMR technologies provide inherent support for group communication, whereas similar features have only recently been included by 3GPP with the introduction of MCPTT in Release 13 [34] and real-time data (MCData) and real-time video (MCVideo) in Release 14 [35]. Table I lists the KPIs for MCPTT service, as suggested by 3GPP [34]. These services leverage GCSE and eMBMS, among others, to optimize radio resources and to perform efficient group management. Carla *et al.* [36] discuss two different approaches in leveraging eMBMS and GCSE to deliver quality services comparable to legacy technologies. The first approach, known as static eMBMS activation, uses multicast mode primarily, whereas unicast mode is configured as a backup. By contrast, the second approach, dynamic eMBMS activation, employs unicast mode only during startup and dynamically switch to multicast when a certain number of group members is reached. Consequently, the second approach is more flexible and spectrally efficient than the first one, owing to its ability to select the most efficient mode. However, Carla *et al.* [36] only consider the centralized approach for the management of group communication. Awada *et al.* [37] review the eMBMS enhancements in Release 13 that include the consideration to employ a single-cell scheme as an enabler for group communication services. Given that eMBMS uses unicast mode for its uplink transmission, the authors propose a new configuration for the single-cell scheme to reduce uplink signaling overhead. In comparison to the existing one that uses the channel quality indicator (CQI) feedback and the acknowledgment of received packets from user devices, the proposed configuration relies only on CQI values. The simulation results show that its performance is similar to the existing configuration while improving the signaling overhead. However, the proposed configuration can only outperform the spectral efficiency of the single-frequency network when the number of user devices per cell is small, thus limiting its use cases. In brief, eMBMS and GCSE enable efficient support for group communication in 4G, owing to their multicast transmissions and flexible group management frameworks.

B. High Reliability

Public safety networks must provide extensive coverage with high reliability due to the critical nature of public safety operations. Ensuring that the critical users have continuous access to communications with the adequate quality of services (QoS) in all scenarios, including during emergencies when the network is usually congested, is crucial. In 4G, the transmission path of a user traffic is known as a bearer. Doumi *et al.* [2] argue that to guarantee high service availability, the network must support bearer prioritization among different services and traffic prioritization among different bearers in dynamic environments. Hence, 3GPP introduces a policy control functionality, which includes QoS class identifier and allocation retention priority parameters, in Release 8 [38]. The former manages prioritization among different bearer traffics during the queuing and scheduling process, whereas the latter is used to define the admission and preemption characteristic of a bearer during congestion. In summary, 4G provides essential features to guarantee adequate QoS for public safety users and services through bearer and traffic prioritizations.

Legacy narrowband PMR systems are proven technologies for public safety-grade communication services. Therefore, 4G must achieve comparable performance to satisfy the requirements of demanding users. Choi *et al.* [39] conduct a feasibility study of MCPTT service deployment on a 4G testbed that employs an NFV-based core network. All components are located at the same site and the testbed configuration is optimized to deliver appropriate policy control with the focus of meeting MCPTT KPIs in Table I. In a simple scenario where only a single group call is active, KPIs 1, 2, and 4b are measured at 162, 420, and 162 ms, respectively, with a probability of 95%. Although the results indicate that the 4G testbed can meet MCPTT delay requirements, Choi *et al.* [39] do not consider practical scenarios, such as those during congestion. In this scenario, Jarwan *et al.* [41] propose the use of an intelligent policy control scheme that is enabled by SDN technology to provide the dynamic prioritization of critical traffic and, subsequently, to satisfy the KPI requirements of MCPTT. Sanchoyerto *et al.* [40] analyze the impact of network evolution from 4G to 5G on MCPTT KPI compliance using a 4G testbed. Multiaccess edge computing (MEC), which is standardized by ETSI to enable the deployment of cloud computing at the edge of a mobile network, is also considered in this article. For all network configurations, the experimental results show that the KPIs are satisfied only in a normal scenario. During congestion, the deployment of distributed EPC on MEC improves the performance of KPIs 1 and 3, where they are measured at 287 and 304 ms, respectively. Although the use of 5G RAN enhances the performance in a normal scenario, Sanchoyerto *et al.* [40] do not consider a congested scenario for this nonstandalone 5G configuration. Briefly, the reliability of the 4G network in guaranteeing adequate QoS, especially during congestion, can greatly benefit from novel features, such as MEC, SDN, NFV, and 5G RAN. Moreover,

5G network slicing that enables the creation of logical networks with appropriate resources may provide a more sophisticated option than 4G policy control in satisfying the KPI requirements of mission-critical services.

C. High Resiliency

Public safety networks must ensure the continuity of communication services in all scenarios, such as during disaster, when the network infrastructure is frequently degraded or destroyed. Jarwan *et al.* [41] state the requirement for solutions similar to rapid deployable network and virtualized network entities to improve resiliency against infrastructure failure. The authors also highlight the need for an efficient sharing of the licensed and unlicensed spectra to enhance network capacity during disaster scenarios wherein cognitive radio technology is suggested as a solution. In addition, Abdul-Ghafoor *et al.* [42] discuss the application of game-theory-based approaches in allocating radio resources among users in a situation where an ad hoc cognitive radio network is accessing a spectrum that is licensed to another operator. Moreover, Doumi *et al.* [2] argue that a highly resilient network can be achieved through careful capacity planning, redundant infrastructure design, and support for direct communication among users in case of partial or complete network failure. To this end, 3GPP introduces the proximity services (ProSe) and isolated E-UTRAN operation for public safety (IOPS) features in Release 12. ProSe allows closely located devices to communicate directly with each other, either with assistance from the network or completely ad hoc. IOPS enables a single or a group of interconnected cells to restore communication services while being isolated from the core network. In brief, these features improve 4G network resiliency against infrastructure failure and facilitate the provisioning of public safety-grade communication services.

Legacy narrowband PMR systems have intrinsic resiliency features, such as base station fallback mode and direct-mode operation in TETRA, which respectively permit an isolated operation of a base station and direct communications between the user devices. The IOPS and ProSe features of LTE are expected to provide similar capabilities but with smaller radio coverage than PMR due to its higher frequency band. Fodor *et al.* [43] highlight the need to leverage the network-assisted and ad hoc modes of ProSe as a key challenge of its implementation in a public safety network. The authors propose a concept that consists of dynamically clustering out-of-coverage devices and assigning the base station role to one of them, known as a cluster head. In the presence of network coverage, the concept further allows the cluster to be integrated into such networks. Three clustering algorithms are considered, and their performance is measured in terms of the percentage of successfully connected devices, the average time for cluster head discovery, and the required energy for the procedure. Fodor *et al.* [43] also believe that the subject requires further research to ensure the viability of mission-critical services over 4G. Furthermore, Favraud *et al.* [44] look into two 4G solutions for dynamic discovery and communication between the mobile and fix base stations with limited or absent backhaul links. The first solution utilizes an evolved user device that can interconnect nearby base stations by associating itself to each of them simultaneously. The second solution utilizes an enhanced base station that is composed of a standard base station, a core network control plane, a local controller, and several virtual user devices. The enhancement enables the base station to operate in the stand-alone mode as proposed by 3GPP IOPS and to form a mesh network with nearby base stations by leveraging its virtual devices for base station discovery and connectivity. As a result, interrupted communication among several base stations can be restored on the fly. The impact of these solutions on link latency between two base stations is studied through experimentations on a 4G network prototype. The maximum delay for the first solution is measured at 68 ms, which decreases as the number of user devices increases. For the second solution, the delay depends on the number of available uplink and downlink subframes, where it is measured at below 100 ms when a single uplink subframe is used. In summary, 4G resiliency can be further improved with intelligent schemes to automate and optimize the operations of IOPS and ProSe. Moreover, 5G network slicing that enables the creation of logical networks with optimized topology may provide greater flexibility in the composition and placement of network functions, which can be leveraged to mitigate the infrastructure failure problems.

D. Comprehensive Security

Public safety users also rely on secure communications due to the sensitive nature of the information carried over the network. McGee *et al.* [45] highlight the security requirements of public safety networks that include the need for data authentication and acknowledgment to ensure the legitimacy of communications. In addition, end-to-end security through data encryption and mutual authentication among different network nodes are also required. Finally, secure access to databases with proper restriction and authorization of different users is a must to avoid data leakage. McGee *et al*. [45] recommend adequate security design from the beginning with periodic assessment as the most effective approach in providing a secure network. Zou *et al.* [46] discuss existing security protocols and algorithms in different wireless networks, including 4G. The latter employs the evolved packet system - authentication and key agreement (EPS-AKA) protocol to provide mutual authentication between the user devices and the core network. The protocol is also responsible for producing keys that are used by 4G cipher known as SNOW3G for data encryption and integrity check. Zou *et al.* [46] also review existing solutions to enhance security at the physical layer for protection against eavesdropping and jamming attacks. Furthermore, the authors highlight the need to jointly optimize the network security and reliability to provide combined protection for different protocol layers and to enable the physical layer security for future 5G radio technologies. Briefly, 4G provides comprehensive and multilayer security features that are essential for the public safety network. Moreover, 5G network slicing that enables the logical isolation of physical resources between different tenants may provide a more secure option to mission-critical services than the 4G network sharing approach.

TABLE II
COMPARATIVE ANALYSIS

References	Scheme	Characteristic	Objective	Strategy
Bega et al. [47]	Automated admission control of network slice	Reliability	Maximize revenue of infrastructure provider while guaranteeing throughput requirement of each slice	Formulate the problem as Semi-MDP and propose a Q-learning algorithm to determine the optimal admission policy.
Bega et al. [48]	Automated admission control of network slice	Reliability	Maximize revenue of infrastructure provider while guaranteeing throughput requirement of each slice	Formulate the problem as Semi-MDP and propose a deep reinforcement learning algorithm to determine the optimal admission policy.
Han et al. [49]	Automated admission control of network slice	Reliability	Maximize long-term network utility that can be defined as revenue of infrastructure provider or QoS level of slice	Propose a genetic algorithm to determine the optimal admission policy.
Aijaz [51]	Agile radio resource slicing	Reliability	Maximize resource utilization while respecting the requirement of each slice type	Formulate the problem as semi-MDP and propose a post-decision state learning algorithm to determine the optimal slicing strategy at the inter-slice level. Then, model the problem as binary integer programming and propose a greedy heuristic algorithm to optimize radio resource allocation at the intra-slice level of haptic communications slice.
Albonda and Perez-Romero [52]	Agile radio resource slicing	Reliability	Maximize resource utilization while respecting the requirement of each slice type	Propose an offline Q-learning algorithm to determine intermediate slicing ratios between slices at the inter-slice level. Then, propose a heuristic-based optimization algorithm to refine the slicing ratio of each slice on the basis of its actual resource requirements.
Li et al. [54]	Secure core network slicing	Security	Maximize slice acceptance ratio while satisfying its resource and security constraints.	Formulate the problem as ILP and propose a heuristic algorithm based on multi-criterion ranking method and k shortest path to optimize the mapping of logical nodes and links respectively.
Sattar and Matrawy [53]	Secure core network slicing	Security and resiliency	Maximize resource utilization while guaranteeing slice delay and isolation requirements.	Formulate the problem as Mixed-ILP with security and delay constraints and propose the simplex method to optimize the mapping of VNF and virtual machine into physical nodes in a small testbed.
Mijumbi et al. [55]	Automated network service reconfiguration	Reliability	Minimize the network service flow time	Propose a heuristic algorithm based on Tabu Search to optimize the VNF mapping into logical nodes and their process scheduling on each of them.
Luo et al. [56]	Automated network service reconfiguration	Reliability	Minimize VNF chain operational cost	Formulate the problem as MDP, then propose a recurrent neural network to predict the rate of traffic flow of a VNF chain and a deep reinforcement learning algorithm to determine the optimal VNF placement in the chain.
Eramo et al. [57]	Automated network service reconfiguration	Reliability and resiliency	Minimize bandwidth rejection and operational cost	Propose a heuristic algorithm to optimize VNF placement in a chain during peak traffic and an MDP-based algorithm to determine VNF optimal migration policy when traffic change occurs.
Tang et al. [58]	Automated network service reconfiguration	Resiliency	Maximize bandwidth utilization and minimize migration cost	Propose an algorithm based on Deep Belief Network to predict future resource requirements. Then, formulate the problem as ILP and propose greedy heuristic and Tabu Search-based algorithms to optimize VNF chain migration by leveraging prediction results.

IV. INTELLIGENT SCHEMES IN 5G NETWORK SLICING FOR PUBLIC SAFETY USE CASE

Section IV reviews the recent studies that propose intelligent schemes for 5G functionalities, as discussed in Section II, which are essential for the implementation of the resource elasticity concept. These schemes leverage artificial intelligence and machine learning methods to meet diverse service requirements through the automated management and agile orchestration of heterogeneous resources at various stages of the network slice lifecycle. Their performance is discussed with regards to public safety network characteristics, as detailed in Section III, especially in improving the reliability and resiliency of network slice to satisfy latency, throughput, and security requirements of mission-critical services. Table II presents a comparative analysis of these schemes in terms of characteristics, objectives, and strategy.

A. Automated Admission Control of Network Slice

Due to the stochastic nature of requests for network slice creation, intelligent resource management is essential to ensure efficient utilization of diverse resources in a multitenant and multiservice environment. One of the approaches is to implement automated network slice admission and allocation control, which is performed before a slice is instantiated. This scheme uses an optimal policy to accept or reject each incoming slice request on the basis of the current system states. Bega *et al.* [47] propose such a scheme in 5G RAN domain to maximize an infrastructure provider's revenue while guaranteeing its tenants' data throughput requirements. System states are defined by the current network load and the required service characteristics. The authors formulate the problem as a semi-MDP and propose an algorithm based on the Q-learning framework to determine the optimal policy. Simulations are performed to compare the

scheme's performance to random policies that randomly reject slice requests and naïve policies that either always accept or reject them. Results show that the scheme provides around 20% and up to 100% gains in revenue over random and naïve policies, respectively, while satisfying the throughput requirements of all slices. However, the scheme's slow convergence to optimal policy limits its scalability in supporting environments with large state space, that is, with a large number of resources and service types. To this end, Bega *et al.* [48] enhance the scheme with a deep reinforcement learning framework, which employs neural networks to generalize experiences that have been learned from observed states. This knowledge is then leveraged to determine the right action for unobserved states with similar features. The experimental results indicate that the framework outperforms naïve policies by up to 100% and around 20%, respectively, in terms of revenue maximization. It also provides fast convergence to optimal policy for small and large system states. Han *et al.* [49] propose a genetic algorithm-based scheme to determine the optimal policy for maximizing long-term network utility under resource constraints. The utility efficiency can be flexibly defined as the revenue of an operator or the QoS level of a slice, while the system states include available resources and types of requested service. Each possible sequence of actions to arriving slice requests is mapped onto an individual binary sequence, known as a strategy. The experimental results reveal that the scheme outperforms static strategies by more than 90% in terms of the solution effectiveness with regards to long-term average network utility. Similar to neural networks, the genetic algorithm also provides inherent support for parallelism, which makes it suitable for complex problems [50]. However, the algorithm's procedure is time-consuming, and it does not guarantee convergence to a global optimum. Although all the proposed admission control schemes improve the efficiency of resource allocations, they focus only on maximizing operators' revenue or utility efficiency. In our opinion, for public safety use cases, the scheme must include slice prioritization while considering penalties that may be imposed when rejecting public safety network slices.

B. Agile Radio Resource Slicing

One of the network slicing challenges in the RAN domain is the need for agile and efficient allocation of scarce resources, such as radio spectrum due to varying demands and channel conditions in relatively short timescales. To this end, researchers propose a two-level RAN slicing scheme wherein radio resources are abstracted and logically controlled using the SDN approach. The objective is to maximize resource utilization while respecting the requirement of each slice type. The scheme combines the multiobjective resource orchestration and slice-aware elastic resource management functions proposed by 5G-MONARCH, where the former optimizes resource allocation across all slices, while the latter performs the task within a slice. Aijaz [51] proposes a slicing scheme that includes haptic communications requiring ultrareliable and low-latency slice type with tight coupling between the downlink and uplink sessions. At the interslice level, the problem is formulated as semi-MDP and is solved using a postdecision state learning algorithm. The algorithm improves the convergence rate of Q-leaning, on which it is based, by exploiting the partial knowledge of the environmental model. At the intraslice level of haptic communications slice, the problem is modeled as a binary integer programming problem and is solved using a greedy heuristic algorithm. A utility function that considers both slice rate and delay is adopted for this slice to satisfy its end-to-end latency requirement between 1 and 2 ms. In comparison to the baseline network-wide slicing approach, the simulation results show that the scheme improves resource utilization by 16% and 18% during normal and high traffic load, respectively. Additionally, the scheme outperforms the baseline approach by 80% in maximizing the sum of utilities, thanks to its ability to dynamically adjust the slice size according to the traffic load. Albonda and Perez-Romero [52] propose a slicing scheme with a focus on mobile broadband and vehicle-to-everything slice types. Similar to the public safety use case, the latter requires low latency and direct communications among vehicles. The scheme first employs an offline Q-learning algorithm to determine the certain intermediate slicing ratios among slices by using an environment model that simulates the network behavior. Then, an online and heuristic-based optimization algorithm refines the ratios on the basis of the actual resource demands of each slice. The objective is to maximize resource utilization while respecting the requirement of each slice type. In comparison to a base scheme that assigns resources proportionally to each slice according to its traffic load, the simulation results show that the proposed scheme achieves a relative gain of 32% for uplink and 20% for downlink in terms of overall resource utilization. The scheme also reduces the access and transmission delays of packets belonging to vehicle-to-everything service from 310 to 130 ms on average. Although all the proposed schemes greatly improve both the latency and resource utilization, they must also consider the security aspect that can be enhanced using the resource isolation feature of network slicing. Furthermore, the agility of RAN slicing schemes can be further improved with the proposed protocol stack redesign by 5G-MONARCH, which reduces interdependencies among network functions.

C. Secure Core Network Slicing

Network slicing allows the logical isolation of physical resources that are shared among slices in such a way that the degradation of a slice does not affect the others. This isolation can be imposed during the network slice instantiation and configuration phase wherein all required resources are created and configured. Researchers propose to leverage this isolation in mitigating security-related issues, such as the distributed denial of service (DDoS) attacks [53] and eavesdropping [54]. These schemes exploit the multiobjective resource orchestration function proposed by 5G-MONARCH to optimize resource allocation across all slices based on multiple objectives, such as minimizing security risks and latency. Li *et al.* [54] propose a core network slicing scheme in an efficient and secure manner to maximize the slice acceptance ratio while satisfying the resource and security constraints. The scheme maps logical nodes and links with specific security requirements onto physical ones with

security levels that meet such requirements. The problem is formulated as an ILP and is solved using a heuristic algorithm. The logical nodes are first provisioned on the basis of a multicriterion ranking method wherein the physical nodes are ranked according to their processing and bandwidth resources and their local and global topologies. Then, the logical links are provisioned using k shortest path method combined with a path selection strategy that considers the bandwidth utilization and hop count of each link. The simulation results show that the scheme improves the slice acceptance ratio compared with existing mapping algorithms not only in normal scenarios but also in the ones where slices have different security requirements. Sattar and Matrawy [53] propose a secure core network slicing scheme to improve slice resiliency against DDoS attacks that involve the starving of a target system resources by sending a large amount of traffic. As mitigations, the scheme enables interslice resource isolation by restricting each physical node to host only one network slice. In addition, it allows intraslice resource isolation wherein different VNFs of a slice are placed on different physical nodes, thus reducing the impact of DDoS attacks. The scheme's objective is to maximize resource utilization while guaranteeing the slice delay and isolation requirements. The authors formulate this optimization problem as mixed ILP and use the simplex method as a solution. The performance of the scheme is evaluated against a greedy algorithm that allocates resources without isolation and on a first-come first-serve basis. The experimental results in a small network testbed show that the scheme reduces the response and round-trip delays associated with the authentication process between the user clients and core network from 150 and 10 ms on average, respectively, below 50 and 1 ms. However, the scheme considers only static resource allocation, which reduces its utilization efficiency. All the proposed core network slicing schemes satisfy the security requirements of a network slice. In our opinion, they can be further improved with automated slice reconfiguration to dynamically mitigate the security issues during the slice runtime phase.

D. Automated Network Service Reconfiguration

A network slice for public safety use cases must guarantee its performance in all scenarios, including during emergencies when traffic is expected to increase significantly. Hence, the slice constituent VNF chains that provide the network services must be reconfigured or scaled throughout their runtime phase, which involves online mapping and scheduling of VNF [55], creating additional VNF instances [56], or increasing the processing capacity of existing ones [57]. These schemes extend the VNF scaling function proposed by 5G-MONARCH by including the reconfiguration and migration of VNF chains. Mijumbi *et al.* [55] propose an efficient scheme of online VNF mapping into logical nodes and their process scheduling on each of them. The authors employ an algorithm based on the tabu search, which is a metaheuristic search method for solving optimization problems. The method iteratively improves an initial random solution by searching for an improved one in its neighborhood. The objective is to minimize the network service flow time that includes the processing and buffering time of all of its constituent VNFs. For evaluation purposes, three greedy optimization algorithms are proposed; each is based either on processing capacity, waiting period, or the current load of candidate nodes. The simulation results indicate that the scheme outperforms the capacity-based greedy algorithm in minimizing the average flow time of network service by nearly 50%. However, the links delay between the physical and virtual nodes is not considered in this scheme. Luo *et al.* [56] propose an agile scaling scheme of a geographically distributed VNF chain to reduce its system cost. This cost includes the overall deployment and operational costs of all VNFs of the chain, as well as the overall flow transfer and delay costs of all flows of the chain. The scheme employs recurrent neural networks combined with a deep reinforcement learning in which the former predicts the rates of traffic flow in a VNF chain, whereas the latter exploits the prediction to decide the optimal placement of its constituent VNFs in terms of their number of instances and deployment locations. The overall delay cost of a chain is obtained by multiplying the end-to-end delay of its flows with a cost per unit of delay. The simulations are performed for different values of the latter to analyze the adaptability of the scheme toward different delay requirements of network services. The results reveal that it outperforms greedy, deep Q networks, and linear-programming-based algorithms by 10% to 37% in reducing the system cost. Although all the proposed schemes improve the operational cost and network revenue, we believe that for public safety network slice, the scheme must focus on guaranteeing slice performance against varying traffic demands and on improving slice resiliency through the allocation of redundant resources.

The reconfiguration and scaling of a VNF chain are limited to the maximum capacity of the physical nodes where its VNFs and virtual links are mapped. Hence, VNF chain migration to new nodes that can satisfy its future resource requirement is also crucial in improving slice resiliency. Eramo *et al.* [57] propose an automated VNF chain reconfiguration and migration scheme with the objective to minimize the bandwidth rejection and operational cost. The latter includes energy consumption and revenue loss due to QoS degradation. Initially, the scheme optimizes VNF chain mapping based on its traffic demand during peak hours. A heuristic-based algorithm is used to decide whether to distribute the placement of its constituent VNFs over a set of servers or to consolidate them based on the average utilization efficiency of both physical servers and links. The consolidation prevents service blocking due to the shortage of link bandwidth. When traffic decreases, the scheme scales the VNF chain in terms of the processing capacity of its constituent VNFs and further consolidates them through the migration process. It uses an MDP-based algorithm to determine optimal migration policy that maximizes energy saving and minimizes QoS degradation caused by the migration. The simulation results indicate that the scheme outperforms a simple policy that does not consider future revenue loss by 27% in minimizing the operational cost. Tang *et al.* [58] propose a scheme for VNF chain migration, which leverages the prediction of its future resource requirements. The objective of the scheme is to minimize system overhead that includes bandwidth and migration overheads. The former reflects the transmission delay of the links, while the latter is

related to the time required for migration. An algorithm based on the deep belief network is used for prediction, which is composed of the offline training of a model using the historical data of resources and the online optimization of prediction parameters using recent data. For VNF chain migration, Tang *et al.* [58] formulate the problem as ILP and propose two algorithms, namely, greedy heuristic and tabu search based. The simulation results show that in minimizing the system overhead accumulated over 120 h, the tabu-search-based scheme outperforms a heuristic algorithm that only considers bandwidth overhead by 32%. All the proposed migration schemes aim at improving the operational costs. In our opinion, for a public safety network slice, the scheme must concentrate on improving the slice resiliency against infrastructure degradation during disaster scenarios and consider the impact on security when migrating the VNF chain to new nodes.

V. Conclusion

In this article, we present a review of the network slicing conceptual framework and its enabling technologies. The concept leverages network softwarization and virtualization technologies to enable the 5G system to support different use cases with diverse characteristics on a common infrastructure by using a logical network approach. Moreover, we examine the characteristics of a public safety broadband network and discuss existing 4G solutions to meet such requirements. In essence, public safety users require group-based communication services deployed on a reliable and resilient network with comprehensive security features. While 4G can meet these requirements during normal scenarios, its performance during emergency and disaster scenarios, wherein network infrastructure is frequently degraded or destroyed, can further be improved with intelligent optimization schemes and emerging technologies, such as NFV, SDN, MEC, and 5G RAN. Furthermore, we review recent studies that propose intelligent schemes for several 5G functionalities that are essential for the efficient and agile management of diverse resources to meet the specific requirements of a customized network slice in the context of public safety use cases. These schemes, namely, intelligent admission control, agile radio resource slicing, secure core network slicing, and automated optimization of network slice, operate at various stages of network slice lifecycle and in different network domains. Although all the proposed schemes meet their target objectives, we believe that for public safety use cases, great efforts should be given to guarantee the continuity of mission-critical services with adequate performance in all scenarios. In conclusion, substantial research efforts are necessary to improve these functionalities to meet the rigid requirements of public safety use cases in terms of network slice reliability, resiliency, and security, especially during the emergency and disaster scenarios.

References

[1] G. Baldini, S. Karanasios, D. Allen, and F. Vergari, "Survey of wireless communication technologies for public safety," *IEEE Commun. Surv. Tut.*, vol. 16, no. 2, pp. 619–641, Apr./Jun. 2014.

[2] T. Doumi *et al.*, "LTE for public safety networks," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 106–112, Feb. 2013.

[3] G. A. Abed, M. Ismail, and K. Jumari, "The evolution to 4G cellular systems: Architecture and key features of LTE-advanced networks," *Int. J. Comput. Netw. Wireless Commun.*, vol. 2, pp. 2250–3501, 2012.

[4] TCCA, "A discussion on the use of commercial and dedicated networks for delivering mission critical mobile broadband services," White Paper Issue 1.2," Feb. 2017.

[5] A. S. A. Latef and R. Hassan, "Spectrum management system: A study," in *Proc. Int. Conf. Elect. Eng. Inform.*, 2011, pp. 1–6.

[6] A. P. Avramova, S. Ruepp, and L. Dittmann, "Towards future broadband public safety systems: Current issues and future directions," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, 2015, pp. 74–79.

[7] "Public-safety broadband network: FirstNet should strengthen internal controls and evaluate lessons learned," U.S. Govern. Accountability Office, Washington, DC, USA, GAO-15-407, 2015.

[8] J. C. Gallagher, "The first responder network (FirstNet) and next-generation communications for public safety: Issues for congress," Emergency Manage., Overview Issues Congr., pp. 79–124, 2018.

[9] "Study on management and orchestration of network slicing for next generation network (release 15)," 3GPP TR 28.801 V15.1.0, Jan. 2018.

[10] A. Nakao *et al.*, "End-to-end network slicing for 5G mobile networks," *J. Inf. Process.*, vol. 25, pp. 153–163, 2017.

[11] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 80–87, May 2017.

[12] P. Rost *et al.*, "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 72–79, May 2017.

[13] A. Othman and N. A. Nayan, "Efficient admission control and resource allocation mechanisms for public safety communications over 5G network slice," *Telecommun. Syst.*, vol. 72, no. 4, pp. 595–607, 2019.

[14] NGMN Alliance, "5G White Paper," V1.0, Feb. 2015.

[15] NGMN Alliance, "Description of network slicing concept," V1.0, Jan. 2016.

[16] X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: Enabling enterprises' own software-defined cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 146–153, Jul. 2016.

[17] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 90–97, Feb. 2015.

[18] ETSI, "Network functions virtualisation—Introductory white paper." [Online]. Available: http://portal.etsi.org/NFV/NFV_White_Paper.pdf, Accessed on: Dec. 2019.

[19] "Network functions virtualisation (NFV); architectural framework," ETSI GS NFV 002 V1.1.1, Oct. 2013.

[20] "Network functions virtualisation (NFV); management and orchestration," ETSI GS NFV-MAN 001 v1.1.1, Dec. 2014.

[21] "SDN architecture issue 1," ONF TR-502, Jun. 2014.

[22] "Applying SDN architecture to 5G slicing issue 1," ONF TR-526, Apr. 2016.

[23] "Network functions virtualisation (NFV) release 3; evolution and ecosystem; report on network slicing support with ETSI NFV architecture framework," ETSI GR NFV-EVE 012 V3.1.1, Dec. 2017.

[24] P. Rost *et al.*, "Mobile network architecture evolution toward 5G," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 84–91, May 18, 2016.

[25] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 32–39, Jul. 15, 2016.

[26] "5G novel radio multiservice adaptive network architecture (5G NORMA) deliverable D3.3 5G NORMA network architecture—Final report version 1.0," 5G-NORMA, Oct. 2017.

[27] D. M. Gutierrez-Estevez *et al.*, "Artificial intelligence for elastic management and orchestration of 5G networks," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 134–141, Oct. 2019.

[28] M. Shariat *et al.*, "A flexible network architecture for 5G systems," *Wireless Commun. Mobile Comput.*, vol. 2019, 2019, Art. no. 5264012.

[29] "Technical specification group services and system aspects; system architecture for the 5G system; stage 2 (release 15)," 3GPP TS 23.501 V15.4.0, Dec. 2018.

[30] "5G mobile network architecture deliverable D2.3 final overall architecture version 1.0," 5G-MoNArch, Apr. 2019.

[31] "5G mobile network architecture deliverable D4.2 final design and evaluation of resource elasticity framework version 1.0," 5G-MoNArch, Apr. 2019.

[32] "Technical specification group services and system aspects; Study on architecture enhancements to support group communication system enablers for LTE (GCSE_LTE) (release 12)," 3GPP TR 23.768 V12.1.0, Jun. 2014.
[33] "Technical specification group services and system aspects; multimedia broadcast/multicast service (MBMS); architecture and functional description (release 10)," 3GPP TS 23.246 V10.2.0, Dec. 2011.
[34] "Technical specification group services and system aspects; mission critical push to talk (MCPTT) over LTE; stage 1 (release 13)," 3GPP TS 22.179 V13.0.0, Dec. 2014.
[35] "Technical specification group services and system aspects; mission critical services common requirements (MCCoRe); stage 1 (release 14)," 3GPP TS 22.280 V14.2.0, Dec. 2016.
[36] L. Carlà, R. Fantacci, F. Gei, D. Marabissi, and L. Micciullo, "LTE enhancements for public safety and security communications to support group multimedia communications," *IEEE Netw.*, vol. 30, no. 1, pp. 80–85, Jan./Feb. 2016.
[37] A. Awada, D. Navrátil, and M. Säily, "A study on single-cell point-to-multipoint transmission for public safety communications with eMBMS LTE networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2016, pp. 1–6.
[38] "Technical specification group services and system aspects; policy and charging control architecture (release 15)," 3GPP TS 23.203 V15.5.0, Jun. 2019.
[39] S. W. Choi, Y.-S. Song, W.-Y. Shin, and J. Kim, "A feasibility study on mission-critical push-to-talk: Standards and implementation perspectives," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 81–87, Feb. 2019.
[40] A. Sanchoyerto, R. Solozabal, B. Blanco, and F. Liberal, "Analysis of the impact of the evolution toward 5G architectures on mission critical push-to-talk services," *IEEE Access*, vol. 7, pp. 115052–115061, 2019.
[41] A. Jarwan, A. Sabbah, M. Ibnkahla, and O. Issa, "LTE-based public safety networks: A survey," *IEEE Commun. Surv. Tut.*, vol. 21, no. 2, pp. 1165–1187, Apr./Jun. 2019.
[42] O. B. Abdul-Ghafoor, M. Ismail, R. Nordin, and A. A. El-Saleh, "Resource allocation in spectrum sharing ad-hoc cognitive radio networks based on game theory: An overview," *KSII Trans. Internet Inf. Syst.*, vol. 7, no. 12, pp. 2957–2986, 2013.
[43] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, and N. Brahmi, "Device-to-device communications for national security and public safety," *IEEE Access*, vol. 2, pp. 1510–1520, 2014.
[44] 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.
[45] A. R. McGee, M. Coutière, and M. E. Palamara, "Public safety network security considerations," *Bell Labs Tech. J.*, vol. 17, no. 3, pp. 79–86, 2012.
[46] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.
[47] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, K. Samdanis, and X. Costa-Perez, "Optimising 5G infrastructure markets: The business of network slicing," in *Proc. IEEE INFOCOM IEEE Conf. Comput. Commun.*, 2017, pp. 1–9.
[48] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, and X. Costa-Perez, "A machine learning approach to 5G infrastructure market optimization," *IEEE Trans. Mobile Comput.*, vol. 19, no. 3, pp. 498–512, Mar. 1, 2020.
[49] B. Han, J. Lianghai, and H. D. Schotten, "Slice as an evolutionary service: Genetic optimization for inter-slice resource management in 5G networks," *IEEE Access*, vol. 6, pp. 33137–33147, 2018.
[50] A. Ansari and A. A. Bakar, "A comparative study of three artificial intelligence techniques: Genetic algorithm, neural network, and fuzzy logic, on scheduling problem," presented at the 4th Int. Conf. Artif. Intell. Appl. Eng. Technol., Kota Kinabalu, Malaysia, 2014, pp. 31–36.
[51] A. Aijaz, "Hap-SliceR: A radio resource slicing framework for 5G networks with haptic communications," *IEEE Syst. J.*, vol. 12, no. 3, pp. 2285–2296, Sep. 2018.
[52] H. D. R. Albonda and J. Perez-Romero, "An efficient RAN slicing strategy for a heterogeneous network with eMBB and V2X services," *IEEE Access*, vol. 7, pp. 44771–44782, 2019.
[53] D. Sattar and A. Matrawy, "Towards secure slicing: Using slice isolation to mitigate DDoS attacks on 5G core network slices," in *Proc. 7th Annu. IEEE Conf. Commun. Netw. Secur.*, 2019, pp. 82–90.
[54] X. Li, C. Guo, L. Gupta, and R. Jain, "Efficient and secure 5G core network slice provisioning based on VIKOR approach," *IEEE Access*, vol. 7, pp. 150517–150529, 2019.
[55] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and S. Davy, "Design and evaluation of algorithms for mapping and scheduling of virtual network functions," in *Proc. 1st IEEE Conf. Netw. Softwarization*, 2015, pp. 1–9.
[56] Z. Luo, C. Wu, Z. Li, and W. Zhou, "Scaling geo-distributed network function chains: A prediction and learning framework," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 8, pp. 1838–1850, Aug. 2019.
[57] V. Eramo, E. Miucci, M. Ammar, and F. G. Lavacca, "An approach for service function chain routing and virtual function network instance migration in network function virtualization architectures," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2008–2025, Aug. 2017.
[58] L. Tang, X. He, P. Zhao, G. Zhao, Y. Zhou, and Q. Chen, "Virtual network function migration based on dynamic resource requirements prediction," *IEEE Access*, vol. 7, pp. 112348–112362, 2019.

Anuar Othman received the B.Sc. degree in electrical and electronics engineering from the University of Rouen, Mont-Saint-Aignan, France, in 2001, and the M.Sc. degree in digital multimedia and communication systems from the University of Strathclyde, Glasgow, U.K., in 2008. He is currently working toward the Ph.D. degree in systems engineering with the National University of Malaysia, Bangi, Malaysia, with a focus on critical communications services in 5G.

He has more than 15 years of experience in telecommunications projects for mobile operators and public safety agencies. In recent years, he has been focusing on the convergence of narrowband and broadband mobile communications for business and mission-critical markets.

Nazrul Anuar Nayan received the B.E. degree in information and communication engineering from the University of Tokyo, Tokyo, Japan, in 1998, the M.E. degree in electrical and electronics and the Ph.D. degree in electronics and information systems engineering from Gifu University, Gifu, Japan, in 2008 and 2011, respectively.

Since, 2014–2016 he had been a Postdoctoral Researcher with the Institute of Biomedical Engineering, University of Oxford, Oxford, U.K. His research interests lie in the field of big data in healthcare (biomedical signal processing), digital-integrated circuit design, and wireless communications.