

3GPP SA2 architecture and functions for 5G mobile communication system^{☆,☆☆}

Junseok Kim^a, Dongmyoung Kim^b, Sunghyun Choi^{a,*}

^a Department of ECE and INMC, Seoul National University, Seoul, Republic of Korea

^b Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea

Received 16 January 2017; accepted 14 March 2017

Available online 13 April 2017

Abstract

The on-going development of Fifth Generation (5G) mobile communication technology will be the cornerstone for applying Information and Communication Technology (ICT) to various fields, e.g., smart city, smart home, connected car, etc. The 3rd Generation Partnership Project (3GPP), which has developed the most successful standard technologies in the mobile communication market such as Universal Mobile Telecommunication System (UMTS) and Long Term Evolution (LTE), is currently carrying out the standardization of both 5G access network system and 5G core network system at the same time. Within 3GPP, Service and System Aspects Working Group 2 (SA2) is responsible for identifying the main functions and entities of the network. In December 2016, the 3GPP SA2 group finalized the first phase of study for the architecture and main functions of 5G mobile communication system under the study item of Next Generation system (NextGen). Currently, normative standardization is on-going based on the agreements made in the NextGen Phase 1 study. In this paper, we present the architecture and functions of 5G mobile communication system agreed in the NextGen study.

© 2017 The Korean Institute of Communications Information Sciences. Publishing Services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: 5G; Next Generation system (NextGen); Mobility management; Session management; QoS framework; Network slicing

1. Introduction

Along with the publication of 5G service and performance requirements by International Telecommunication Union-Radio communication (ITU-R) [1], many research groups today are conducting research and standardization activities of 5G mobile communication system actively [2]. 3GPP has specified new use cases that cannot be met with 4G Evolved Packet System (EPS) [3]. The proposed use cases can be classified into five categories according to their examples as follows.

- Enhanced mobile broadband: Ultra High Definition (UHD), virtual presence.
- Critical communication: Robot/drone, emergency.
- Massive machine type communication: eHealth.

- Network operation: Network slicing, interworking.
- Enhancement of vehicle-to-everything: Autonomous driving.

It is necessary to reduce latency and connect many devices to the network while increasing data rate in order to support the above services. 5G mobile communication system will enable not only existing services, but also new services in various fields. Through realization of the new services, Information and Communication Technology (ICT) convergence will come true in real life.

3GPP has been working on the standardization of 5G mobile communication system for the commercialization of 5G in 2020. Radio Access Network (RAN) and Service and System Aspects (SA) are the representative Technical Specification Groups (TSGs) within 3GPP. The TSG RAN is developing documents covering radio access architecture and radio interface protocol aspects of new Radio Access Technology (RAT) [4,5]. SA Working Group 2 (SA2) within the TSG SA studies architecture and main functions of the 5G network system under the study item of Next Generation system (NextGen). The SA2 finalized the NextGen Phase 1 study in December 2016 and published the 3GPP TR 23.799

* Corresponding author.

E-mail addresses: jskim14@mwnl.snu.ac.kr (J. Kim), dm.kim@etri.re.kr (D. Kim), schoi@snu.ac.kr (S. Choi).

Peer review under responsibility of The Korean Institute of Communications Information Sciences.

[☆] This paper is part of a special section titled “Special Section on ICT Convergence Technology”.

^{☆☆} This paper has been handled by Prof. Jungwoo Lee.

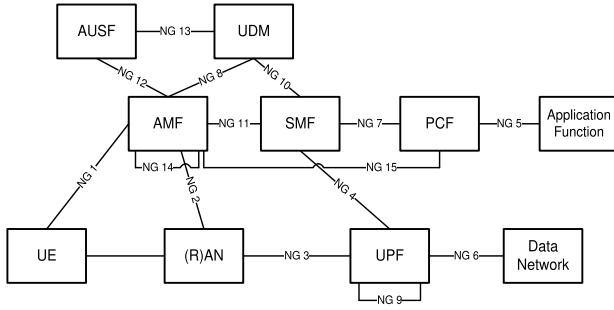


Fig. 1. Architecture and reference points for NextGen.

specification as an outcome of the study [6]. NextGen Phase 2 study is also expected to begin mid 2017. Based on the agreements within the NextGen study, SA2 group is currently conducting normative standardization work for 5G mobile core network architecture by aiming to finalize the initial architecture by the end of 2017. In this paper, we present the architecture and functions of 5G mobile communication system by referring to the 3GPP SA2 NextGen Phase 1 study.

The rest of the paper is structured as follows. The architecture for NextGen based on the agreements is presented in Section 2, and the agreements on various functions of NextGen are detailed in Section 3. Key functions of EPS and NextGen are compared in Section 4. Lastly, Section 5 concludes the paper.

2. Architecture for next generation core network

In 4G core network, called Evolved Packet Core (EPC), protocol and reference points are defined for each entity such as Mobility Management Entity (MME), Serving Gateway (S-GW), and Packet Data Network Gateway (P-GW). On the other hand, in NextGen, protocol and reference points are defined for each Network Function (NF). We now present the NextGen architecture and the agreements of overall architecture.

2.1. NextGen architecture overview

Overview: Fig. 1 shows the NextGen architecture composed of NFs and reference points connecting NFs. User Equipment (UE) is connected to either RAN or Access Network (AN) as well as Access and Mobility Function (AMF). RAN represents a base station using new RAT and evolved LTE while AN is a general base station including non-3GPP access, e.g., Wi-Fi. The NextGen core network consists of various NFs. In Fig. 1, there are seven NextGen core NFs, namely, (1) AMF, (2) Session Management Function (SMF), (3) Policy Control Function (PCF), (4) Application Function (AF), (5) Authentication Server Function (AUSF), (6) User Plane Function (UPF), and (7) User Data Management (UDM).

Network functions: NF, 3GPP-adopted processing function in NextGen, has both functional behavior and interface. An NF can be implemented either as a network element on a dedicated hardware, as a software instance running on a dedicated hardware, or as a virtualized function instantiated on an appropriate platform, e.g., a cloud infrastructure [6].

Functional description: AMF provides UE-based authentication, authorization, mobility management, etc. A UE even using multiple access technologies is basically connected to a single AMF because the AMF is independent of the access technologies. SMF is responsible for session management and allocates IP addresses to UEs. It also selects and controls the UPF for data transfer. If a UE has multiple sessions, different SMFs may be allocated to each session to manage them individually and possibly provide different functionalities per session. AF provides information on the packet flow to PCF responsible for policy control in order to support Quality of Service (QoS). Based on the information, PCF determines policies about mobility and session management to make AMF and SMF operate properly. AUSF stores data for authentication of UE while UDM stores subscription data of UE. Data network, not part of NextGen core network, provides Internet access or operator services.

Reference points: Reference point representation of the architecture can be used to develop detailed call flows in the normative standardization. Next Generation (NG)1 is defined to carry signaling between UE and AMF. The reference points for connecting between AN and AMF and between AN and UPF are defined as NG2 and NG3, respectively. There is no reference point between AN and SMF, but there is a reference point, NG11, between AMF and SMF. Therefore, we can confirm that SMF is controlled by AMF. NG4 is used by SMF and UPF so that the UPF can be set using the control signal generated by the SMF, and the UPF can report its state to the SMF. NG9 is the reference point for the connection between different UPFs, and NG14 is the reference point connecting between different AMFs, respectively. NG15 and NG7 are defined since PCF applies policy to AMF and SMF, respectively. NG12 is required for the AMF to perform authentication of the UE. NG8 and NG10 are defined because the subscription data of UE is required for AMF and SMF.

2.2. Agreements on overall architecture

Separation of control and user planes: NextGen aims at separating user plane and control plane. The user plane carries user traffic while the control plane carries signaling in the network. In Fig. 1, the UPF is in the user plane and all other NFs, i.e., AMF, SMF, PCF, AF, AUSF, and UDM, are in the control plane. Separating the user and control planes guarantees each plane resource to be scaled independently. It also allows UPFs to be deployed separately from control plane functions in a distributed fashion. In this architecture, UPFs may be deployed very close to UEs to shorten the Round Trip Time (RTT) between UEs and data network for some applications requiring low latency.

Modularization: NG architecture is composed of modularized functions. For example, the AMF and SMF are independent functions in the control plane. Separated AMF and SMF allow independent evolution and scaling. Other control plane functions like PCF and AUSF can be separated as shown in Fig. 1. Modularized function design also enables NextGen to support various services flexibly.

Interaction: Each NF interacts with another NF directly. It is not impossible to use an intermediate function to route messages from one NF to another NF. In the control plane, a set of interactions between two NFs is defined as service so that its reuse is possible. This service enables support for modularity. The user plane supports interactions such as forwarding operations between different UPFs.

Roaming: NextGen considers the architecture for supporting roaming in a manner similar to how EPS does. There are two kinds of deployment scenarios, i.e., Home Routed (HR) and Local Break Out (LBO). The two structures to support roaming and the session management associated with them are described in Section 3.2.

3. Agreements on NextGen functions

In the NextGen core network, existing 4G core network technologies are required to be improved in order to efficiently support various services and requirements. In this section, we introduce the agreements on the NextGen functions discussed in 3GPP, namely, (1) mobility management, (2) session management and service continuity, (3) QoS framework, and (4) network slicing.

3.1. Mobility management

State model: Fig. 2 shows the state models of EPS and NextGen. Three types of states are shown in the EPS model, i.e., EPS Mobility Management (EMM) state, EPS Connection Management (ECM) state, and Radio Resource Control (RRC) state [7]. EMM and ECM states are managed by the core network, where the EMM state represents whether a UE is registered in the EPC, and the ECM state shows whether Non Access Stratum (NAS) signaling connection between UE and MME is established. On the other hand, RRC state is managed by RAN, and it represents whether a connection between UE and RAN, i.e., evolved Node B (eNB), exists or not. A UE in the ECM-CONNECTED state needs to be in the RRC-CONNECTED state, because radio link connection is required to establish NAS signaling connection.

In NextGen, the MM state of a UE can be MM-REGISTERED or MM-DEREGISTERED state depending on whether the UE is registered in NextGen core network, which is very similar to EMM-REGISTERED and EMM-DEREGISTERED states [6]. When the UE is registered, the UE is in either Core Network (CN)-IDLE state or CN-CONNECTED state according to the existence of NAS layer connection. The definitions of MM and CN states are almost the same as those of EMM and ECM states in the EPS. On the other hand, RRC-INACTIVE CONNECTED state is newly introduced as a state of the RRC state model. The new state is proposed to be used as a primary sleeping state prior to RRC-IDLE state [8]. When a UE moves to the new state, both the UE and RAN keep the context information of the UE's RRC connection, such as UE capabilities and security context, that have been obtained during the RRC connection

setup. Therefore, the new state enables a lightweight transition from inactive to active data transmission.

Handover and cell reselection: In the EPS, when a UE is in the RRC-CONNECTED state, the serving eNB evaluates the reported signal strength between the UE and the eNB, and performs a handover procedure when the signal strength is weakened. However, in the RRC-IDLE state, where the eNB is not aware of the existence of the UE, the UE decides whether to camp on the current cell or to reselect a neighboring target cell based on signal strength measurements. This procedure is referred to as cell reselection. In the EPS, the utilized mobility procedures are fixed as handover in the RRC-CONNECTED state and cell reselection in the RRC-IDLE state, respectively. On the other hand, in NextGen, it is expected that the core network is able to flexibly control whether to perform handover or cell reselection for a UE in CN-CONNECTED state.

Location tracking: In the EPC, the location of a UE is managed by the MME. The level of a UE's location is different according to the RRC state of the UE. In the RRC-CONNECTED state, the UE's location is tracked in the cell level while in the RRC-IDLE state, its location is tracked in the Tracking Area (TA) level, which is a set of cells. Similarly, in NextGen, the core network can track the location of UE at the CN location area level in the CN-IDLE state, and the UE location is known at the level of the serving RAN to the core network in the CN-CONNECTED state. A CN location area is an area allocated by the network registering the UE, i.e., TA list. The NextGen core network can allocate TA list composed of cells using new RAT and evolved LTE. On the other hand, in NextGen, RAN also needs to support the location tracking for the UE in RRC-INACTIVE CONNECTED state. In that state, the core network understands that the UE is located within the RAN area, but the RAN needs a new location tracking functionality to determine the exact location of the UE because the connection between the UE and the RAN is not active.

Paging: In the EPS, when Down Link (DL) traffic for a UE in the RRC-IDLE state arrives at the S-GW, the MME performs a paging procedure based on the detected location of the UE. On the other hand, it was agreed to support the following two kinds of paging, i.e., CN paging and RAN paging, in the NextGen system. CN paging, the default paging procedure, is requested by the core network when the UE is in the CN-IDLE state. A newly introduced RAN paging is needed for UE in the RRC-INACTIVE CONNECTED state. Since a UE in the RRC-INACTIVE CONNECTED state is in the CN-CONNECTED state in the core network's viewpoint, the core network simply forwards the data or the signaling message to the corresponding RAN when the data or signaling message arrives. Therefore, RAN itself generates the paging message and performs paging to find the exact location of the UE, and then to send the data or signaling message to the UE. The NextGen core network can transmit additional assistance information to RAN for RAN paging.

Mobility on demand: Mobility on demand is a concept to support mobility not to all devices but only to devices that need it [9]. It also includes supporting UE's mobility at its

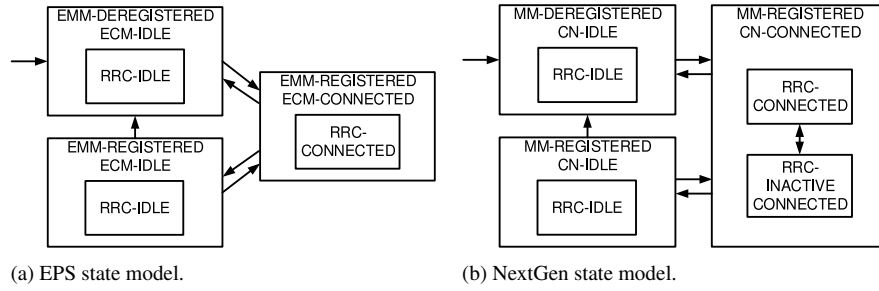


Fig. 2. State model of 4G EPS and 5G NextGen.

appropriate level. There had been many discussions on mobility on demand, and it is divided into aspects of mobility restriction and mobility pattern (or mobility level). The mobility restriction is addressed in terms of area, which is divided into allowed area, non-allowed area, and forbidden area. The granularity of the area is at least TA level. In the allowed area, UE can communicate through the control plane or the user plane. UE cannot send service request and session management signaling in the non-allowed area. However, periodic registration update is possible. It can also respond to the paging of the NextGen core network. Moreover, emergency calls or multimedia priority service are allowed. In the forbidden area, UE is not allowed to have any communication with the network except for the emergency services.

The mobility pattern is used as a concept to describe the expected mobility of UE in the NextGen core network, not a parameter delivered on the interface defined in the standard. The mobility pattern is determined by considering subscription, location, capabilities, and mobility information statistics of UE, network policies, etc. The NextGen core network can use the mobility pattern to optimize the mobility management procedure and related parameters of UE.

Mobile Originated (MO) only mode: Internet of Things (IoT) service is an important 5G service. IoT devices, e.g., sensor devices, mostly send MO data. For this kind of devices, MO only mode is defined in NextGen, and the NextGen core network determines whether to apply the MO only mode to a UE during the registration procedure based on the subscription data of the UE and network policy. The MO only mode is allocated to a UE, which does not require Mobile Terminated (MT) traffic. Therefore, the UE in MO only mode does not listen to the paging message. The NextGen core network does not need to manage the UE's location while it is registered in the NextGen core network. For optimization, the NextGen core network may decide to deregister after the MO data communication is finished, without transferring the UE's state into the CN-IDLE state in the MM-REGISTERED state, because most functions supported in the CN-IDLE state is not meaningful for the UE in MO only mode, e.g., UE location tracking and reachability management. In such cases, the UE needs to perform attach procedure whenever the MO data transmission is necessary to communicate with the core network.

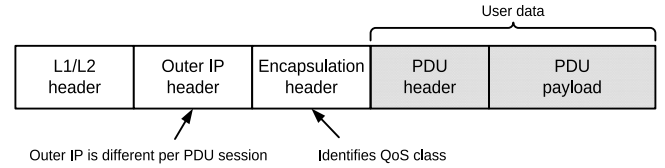


Fig. 3. Packet format for session tunnel protocol.

3.2. Session management and service continuity

PDU session: UE receives services through a Protocol Data Unit (PDU) session, which is a logical connection between the UE and data network. In NextGen, various PDU session types are supported, e.g., IPv4, IPv6, Ethernet, etc. Unlike the EPS, where at least one default session is always created while the UE attaches to the network, NextGen can establish a session when service is needed independently of the attachment procedure of UE, i.e., attachment without any PDU session is possible. NextGen also supports UE establishing multiple PDU sessions to the same data network or to different data networks over a single or multiple access networks including 3GPP and non-3GPP accesses.

The number of UPFs for a PDU session is not specified. At least, deployment with one UPF is essential to serve a given PDU session. For a UE with multiple PDU sessions, there is no need for convergence point like S-GW in the EPC. In other words, the user plane paths of different PDU sessions are completely disjoint. This implies that there is a distinct buffering node per PDU session for the UE in the RRC-IDLE state.

Fig. 3 shows the user plane packet format, which can support tunneling per PDU session. All QoS classes of a session share the same outer IP header, but the encapsulation header carries different QoS markings. The content about QoS marking is introduced in Section 3.3

MM and SM interaction: When a UE attaches to the NextGen core network, both the MM and SM related messages and procedures are exchanged between the UE and the AMF through the NG1 NAS connection. SM NAS message can be transmitted with the MM NAS message supported by the NAS routing capability within the AMF. Although the AMF is involved in sending NAS SM messages, processing of NAS MM and NAS SM messages occurs in the AMF and SMF, respectively.

The AMF can select SMF functions for PDU sessions in addition to aforementioned features in Section 2.1. When

a UE requests a new PDU session, it provides SM service identification information with the request message. Although the AMF does not understand full context of the SM related NAS messages, it has to handle the SM service identification information to select a proper SMF for the new PDU session. The SM service identification information contains the network slicing related information and the Data Network Name (DNN). The AMF uses the SM service identification information together with subscription data of UE from SDM to select proper SMF.

When an SMF is selected for a particular PDU session, the AMF shall ensure that all NAS signaling associated with this PDU session is handled by the same SMF. Therefore, PDU session identifier is assigned. UE uses this identifier to allow the message to be delivered to correct SMF.

Session and service continuity: In the EPS, continuity of IP session for all UEs is guaranteed in the whole system area. That is, the P-GW and the IP address of UE's PDU session are maintained regardless of the location of the UE such that the session continuity is maintained. On the other hand, NextGen aims to provide various levels of session continuity depending on the type of UE and type of service.

Service continuity means uninterrupted user experience of a service, even in the cases of the IP address and/or anchoring point changes [6]. Some applications may not require such service continuity depending on the characteristics of the service. Furthermore, some other applications may not essentially need the guarantee of IP session level continuity even though the service continuity is required, e.g., due to existence of separate functions for providing service continuity at the upper layer. Accordingly, the NextGen system does not always guarantee session continuity, but provides it flexibly.

NextGen tries to use system resources effectively by providing an appropriate level of Session and Service Continuity (SSC) per PDU session. There are three types of SSC modes, namely, SSC mode 1, SSC mode 2, and SSC mode 3. Separate SSC modes are closely related to how the Terminating User Plane Function (TUPF) of PDU session is allocated and managed. The TUPF terminates the user plane in the NextGen core network and interfaces with the data network. It is responsible for IP anchoring as P-GW does in the EPC.

In SSC mode 1, session continuity is guaranteed in all areas by keeping the TUPF identical regardless of access network type and UE location as in the EPS. In SSC mode 2, the same TUPF is maintained across only a subset of the whole network, referred to as the serving area of the TUPF. When a UE leaves the serving area of a TUPF, the UE will be served by a different TUPF suitable for the UE's new point of attachment to the network. Lastly, SSC mode 3 enables the UE to connect to a new TUPF before the connection with the existing TUPF is released, and hence, the UE can receive services from two TUPFs at the same time. By using this mode, various optimization can be supported, e.g., a newly created session is served by a new TUPF located in a preferred location while maintaining the existing sessions in the previous TUPF to provide session continuity.

Roaming: Public Land Mobile Network (PLMN) is a network that is established and operated by a carrier for providing mobile communication services to the subscribers. Domestic carrier's subscribers use roaming if they want to receive services from abroad. When a subscriber's UE of domestic carrier is registered in the abroad network, the previously used network is Home PLMN (HPLMN), and the currently registered network is Visited PLMN (VPLMN).

In case of roaming, there are two possible deployment scenarios for PDU session as mentioned in Section 2.2. Fig. 4 shows the architecture for the HR and LBO scenarios for roaming. In the HR scenario, a PDU session is supported by two SMFs. One SMF is under the control of the HPLMN while the other is under the control of VPLMN. There should be at least one UPF in the HPLMN and at least one UPF in the VPLMN. In case of LBO, SMF and all UPFs involved in a PDU session are under the control of the VPLMN.

3.3. QoS framework

EPS: QoS indicates the performance particularly from the perspective of network users. In the EPS, the QoS is managed based on EPS bearer in the EPC and RAN. That is, all types of traffic mapped to the same EPS bearer receive the same level packet forwarding treatment.

QoS flow: QoS flow is the finest granularity for QoS management in the NextGen core network. A QoS flow can either guarantee the bit rate or not, i.e., Guaranteed Bit Rate (GBR) QoS flow or non-GBR (NGBR) QoS flow. The bearer concept is not considered in the NextGen core network while 5G RAN is expected to maintain the radio bearer concept. By introducing the QoS flow concept, granularity for QoS treatment (per flow) becomes different from tunneling (per PDU session, basically), and it will enable more flexible QoS control.

QoS rule: NextGen defines QoS rule, which consists of QoS profile, packet filters, and precedence order. The QoS profile is composed of QoS parameters and QoS marking. The value of the QoS marking indicates the type of QoS profile, i.e., A-type or B-type. A-type QoS profile has standardized QoS parameters where B-type QoS profile has QoS parameters dynamically signaled over NG2. Packet filter is used for the purpose of binding a flow to a specific QoS marking. Precedence order represents the priority to adapt QoS rule to a flow.

The QoS parameters for the QoS profile type agreed in NextGen study are summarized in Table 1. All parameters except notification control are used in the EPS. The notification control parameter controls whether notification should be made if the QoS targets are no longer fulfilled for QoS flows. Priority level, packet delay budget, and packet error rate apply for both A-type and B-type QoS profiles. The other parameters apply only to the B-type QoS profile. Packet filters in the QoS rule provided by the control plane of the NextGen core network are used for the purpose of binding to specific QoS marking. Precedence order is the parameter, which determines the QoS rule to be applied to a flow.

Reflective QoS: Network decides on the QoS to apply on DL traffic, and UE reflects the DL QoS to the associated UL traffic.

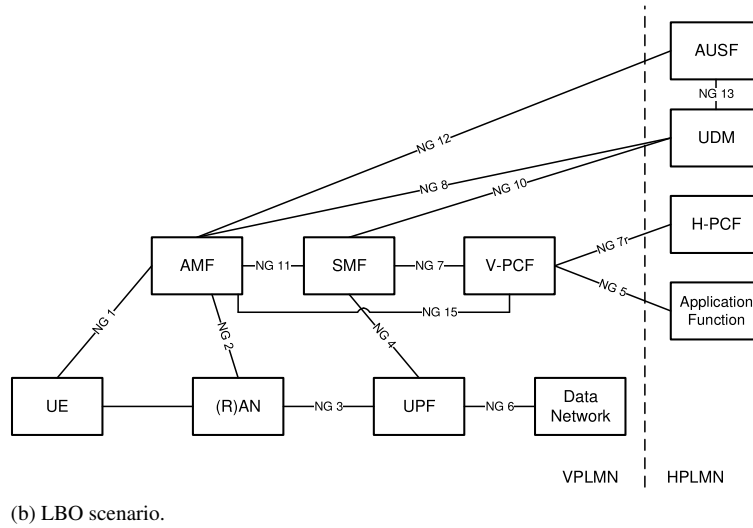
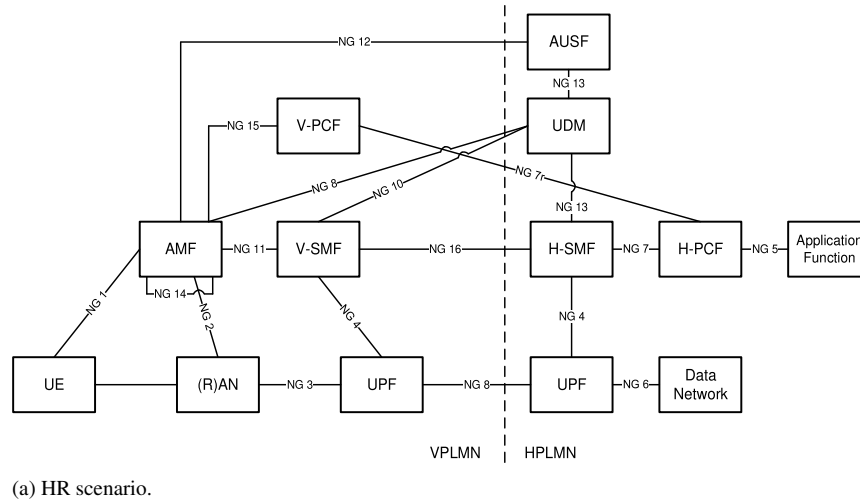


Fig. 4. Roaming deployment scenarios for NextGen.

Table 1
QoS parameters for NextGen.

QoS parameters	QoS profile type	
	A-type	B-type
DL/UL GBR for flow	×	○
DL/UL MBR ^a for flow	×	○
Priority level	○	○
Packet delay budget	○	○
Packet error rate	○	○
ARP ^b	×	○
Notification control	×	○

^a Maximum Bit Rate (MBR) is the maximum number of bits allowed at a given amount of time, usually a second.

^b Allocation and Retention Priority (ARP) is a parameter used to determine whether to release an existing EPS bearer and generate a new EPS bearer or reject the new EPS bearer when there are insufficient resources in the EPS.

When UE receives a DL packet for which reflective QoS should be applied, the UE creates a new derived QoS rule. The packet filter in the derived QoS rule is derived from the DL packet. It is possible to apply both reflective QoS and non-reflective QoS on the same PDU session.

3.4. Network slicing

Concept: The network slicing concept proposed by Next Generation Mobile Networks (NGMN) Alliance consists of three layers as depicted in Fig. 5, namely, (1) service instance layer, (2) network slice instance layer, and (3) resource layer [9]. The service instance layer represents the services supported by the network. Each service is represented by a single service instance. In general, services are provided by network operators or third parties. The network slice instance provides the network characteristics required by the service instance. A single network slice instance can be shared by multiple service instances. The network slice instance is composed of none, one, or more sub-network instances shared by other network slice instances. The sub-network instance is a set of NFs, which run on the physical or logical resources. The network slice is a complete logical network providing telecommunications services and network capabilities. Network slices vary depending on the features of the service they need to support.

Network slicing in NextGen: SA2 has developed high level architecture for supporting network slicing based on its re-

Table 2

Comparison of key functions between 4G EPS and 5G NextGen.

System	Mobility management				Session management		QoS framework QoS granularity
	RRC state	Cell switch	Location tracking level	Paging	Session establishing time	Session continuity	
EPS	RRC-CONNECTED RRC-IDLE	Handover Cell reselection	Cell Tracking area	No Yes	Attachment procedure	Guaranteed in the whole area	EPS bearer
NextGen	RRC-CONNECTED RRC-INACTIVE CONNECTED	Handover Cell reselection	RAN	No RAN paging	Service request procedure	Depending on the type of UE and type of service (SSC mode)	QoS flow
	RRC-IDLE	Cell reselection	CN location area	CN paging			

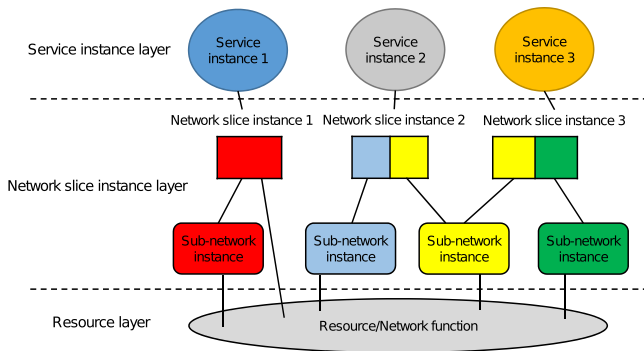


Fig. 5. Components of network slicing.

quirement [3,6]. The network slice in NextGen will provide a customized 5G network service by selecting the control plane and user plane NFs required for the specific service. Accordingly, it is possible to optimize the network for the specific service, thus ensuring the independence and stability of service.

Slice selection: Fig. 6 shows the architecture of NextGen system to support network slicing. Common Control Network Functions (CCNF) include a set of NFs, e.g., AMF and Network Slice Selection Function (NSSF), which should be commonly supported for all the sessions of a specific UE. A UE is connected to and served by a single CCNF at a time though multiple PDU sessions of the UE may be served by different network slice instances. UE level mobility management, authentication, network slice instance selection can be the examples of commonly requested functionalities provided by CCNF.

When network deploys the network slice, it uses Network Slice Selection Assistance Information (NSSAI) provided by UE. UE stores a configured and accepted NSSAI per PLMN. NSSAI is configured by UE to be used in PLMN before any interaction with the network. After the network accepts a UE,

the accepted NSSAI is used by the UE. The attach accept message includes the accepted NSSAI, where it is updated by MM procedures. During the initial attach procedure of the UE, the NSSAI is used by RAN to select the CCNF. NSSF in CCNF selects the network slice instance. Network slice instance selection can be performed based on NSSAI, DNN of the requested session, UE subscription data, and so on.

4. Summary of key functions of EPS and NextGen

Key functions of EPS and NextGen are summarized in Table 2. In mobility management of the NextGen, the biggest difference from the EPS is the addition of the new RRC state, i.e., RRC INACTIVE CONNECTED state. As the new state is added, corresponding cell switch method and location tracking level are defined accordingly, and these parts are not different from the EPS. However, RAN paging is newly defined in the new state. In the EPS, a session is established with the attachment procedure of a UE. In NextGen, however, the session is established when service is needed after the UE attaches. Also, session continuity is not always guaranteed in the NextGen. The QoS flow is the finest granularity for QoS management in the NextGen core network because the bearer concept is not considered.

5. Concluding remarks

In this paper, we presented the agreements on the architecture and functions of NextGen that aims to create a 5G architecture supporting a variety of services, e.g., virtual presence, eHealth, connected car, etc. In the NextGen architecture, NFs are modularized in order to make the network flexible and scal-

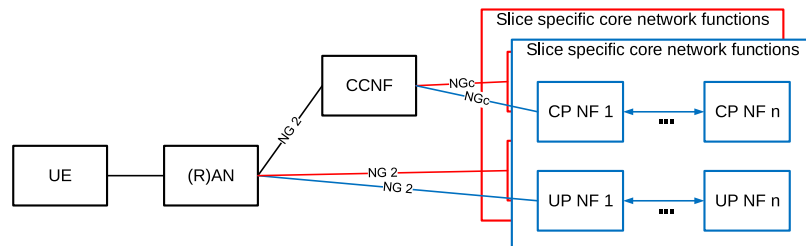


Fig. 6. Network slicing architecture in NextGen.

able. New technology such as network slicing is introduced, and functions such as mobility management, session management, and QoS framework are also developed in different ways from 4G EPS. Detailed procedures about functions as well as normative standardization will be defined in the NextGen Phase 2 study based on the agreements made in Phase 1.

Acknowledgments

This work was supported by ICT R&D program of MSIP/IITP, [B0126-15-1012, Multiple Access Technique with Ultra-Low Latency and High Efficiency for Tactile Internet Services in IoT Environments] and [R7116-16-1001, Development of 5G Core Network Technologies Standards].

Conflict of interest

The authors declare that there is no conflict of interest in this paper.

References

- [1] ITU-R M.2083-0, IMT-Vision-Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond (Sep. 2015).
- [2] [R.N. Mitra, D.P. Agrawal, 5G mobile technology: A survey, *ICT Express* 1 \(3\) \(2016\) 132–137.](#)
- [3] 3GPP TR 22.891, Feasibility Study on New Services and Markets Technology Enablers; Stage 1, V.14.0.0 (Mar. 2016).
- [4] 3GPP TR 38.801, Study on New Radio Access Technology; Radio Access Architecture and Interfaces, V.1.0.0 (Dec. 2016).
- [5] 3GPP TR 38.804, Study on New Radio Access Technology; Radio Interface Protocol Aspects, V.0.4.0 (Nov. 2016).
- [6] 3GPP TR 23.799, Study on Architecture for Next Generation System, V.1.2.1 (Nov. 2016).
- [7] 3GPP TS 23.401, General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access, V.14.2.0 (Dec. 2016).
- [8] I.L. Da Silva, G. Mildh, M. Säily, S. Hailu, A novel state model for 5G radio access networks, in: *Proc. IEEE ICC*, 2016, pp. 632–637.
- [9] NGMN Alliance, 5G White paper, V.1.0 (Feb. 2015).