O-RAN next Generation Research Group (nGRG)

Contributed Research Report

Research Report on Service-based RAN for 6G Network

Report ID: RR-2024-11

Contributors:

CUC

Mavenir

Nokia

Lenovo

CICT

Release date: 2024.10

Authors

Shan Liu, CUC (Editor-in-Chief)

Rong Huang, CUC

Christian Mannweiler, Nokia

Sapna Sangal, Mavenir

Yuanfang Huang, CICT

Congchi Zhang, Lenovo

Makarand Kulkarni, Tejas

Reviewers

Douglas Knisely, Qualcomm Technologies, Inc

Jiaying Zong, China Telecom

Massimo Condoluci, Ericsson

Ravi Sinha, Reliance Jio

Clifton Fernandes, Nokia

Disclaimer

Copyright

Executive summary

The 5G Core Network introduces a Service-based Architecture in the control plane, further integrating the Information Technology concept (e.g., cloud, service-based) into the mobile network, and bringing more flexibility and convenience for service evolution. As we look forward to 6G, more use cases and new capabilities will be provided. The evolution of RAN architecture has also become inevitable, especially for the cloud-native RAN. The evolving SBA into the RAN architecture for 6G is a valuable research topic for investigation.

This research report analyzes the motivations and requirements for service-based RAN, which include the customized needs of industry users, capacity exposure, the introduction of new capabilities (AI, sensing) for 6G, service-oriented network management/orchestration and business innovation. Next, it introduces the use cases of service-based RAN, e.g., industry-specific customized networks, edge convergence networks, emergency communication, etc. For the basic considerations of Service-based RAN, this report analyzes the principles to be followed, including modular service definition, service reusability, self-contained functional structuring, etc., and analyzes the different directions of service implementation from different architectural perspectives such as control plane, user plane, and service access interfaces.

Based on the principles and basic directions, this research report gives two different Service-based RAN solutions: solution 1 is interface-based, which is more aligned with the existing 5G architecture; solution 2 is more closely following the concepts of SBA, where all RAN functions are restructured, and it is easier to introduce new AI, computing and other capabilities. In this case, it is potentially more conducive to taking full advantages of cloud-native design principles. Finally, the challenges faced by Service-based RAN are analyzed, including the rationality of service definition, performance impact, UE impact, interoperability, etc.

Service-based RAN is an important direction for the evolution of RAN in the future. This report gives a preliminary study and expects industry to promote the maturity of technical solutions, enabling more scenarios and better networks for 6G.

Table of Contents

List of abbreviations

3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
AMF	Authentication Management Function
API	Application Programming Interface
CN	Core Network
CU	Central Unit
CU-CP	Central Unit Control Plane
CU-UP	Central Unit User Plane
DU	Distributed Unit
GTP	General Transport Protocol
HTTP	Hyper Text Transfer Protocol
JSON	JavaScript Object Notation
MEC	Mobile Edge Computing
MSA	Micro-Service Architecture
NAS	Non-Access Stratum
NRF	Network Repository Function
NF	Network Function
NFS	Network Function Service
P2P	Point-to-Point
PCF	Policy Control Function
PDCP	Packet Data Convergence Protocol
QUIC	Quick UDP Internet Connections
QoS	Quality of Service
RLC	Radio Link Control
SCP	Service Communication Proxy
SDAP	Service Data Adaptation Protocol
SBA	Service-Based Architecture
SBI	Service-Based Interface
TCP	Transmission Control Protocol
UDSF	Unstructured Data Storage Function

UPF	User Plane Protocol

List of figures

1 Background

1.1 Introduction of SBA in 5G

The continuous evolution of Information Technology (IT) brings rapid innovation in multiple fields such as the Internet. At the same time, higher demands for mobile communications are arising for more industries and use cases. This drives the need for evolution and innovation in mobile networks. The integration of Information and Communications Technology is gradually becoming an essential trend.

Traditional 3GPP Core Network (CN) functions are more centralized and usually highly integrated with dedicated hardware. In the 5G era, traditional designs for the CN were abandoned, and an innovative Service-Based Architecture (SBA) for Control Plane (CP) was introduced in 3GPP[1]. SBA in 5G is a system architecture in which the system functionality is achieved by a set of Network Functions (NFs) providing services to other authorized NFs to access their services. It is unlike the older architectures in which system functionalities are achieved by procedures defined as part of Point-to-Point (P2P) interfaces between two network nodes.

SBA has commonalities with the "Micro-Service Architecture (MSA)" in Cloud-Native, which decouples monolithic and integrated software into small, independent units, thereby enabling rapid iteration of Micro-Service based functions to cope with the rapid growth of the Internet[2]. The 5G CN rebuilds traditional network elements into different functional modules called NFs (the subsequent service-based RAN NFs in this RR are also based on this concept). A NF can contain multiple Network Function Services (NFS). Thus, the 5G CN can be characterized as cloud native.

SBA in 5G CN is shown below.

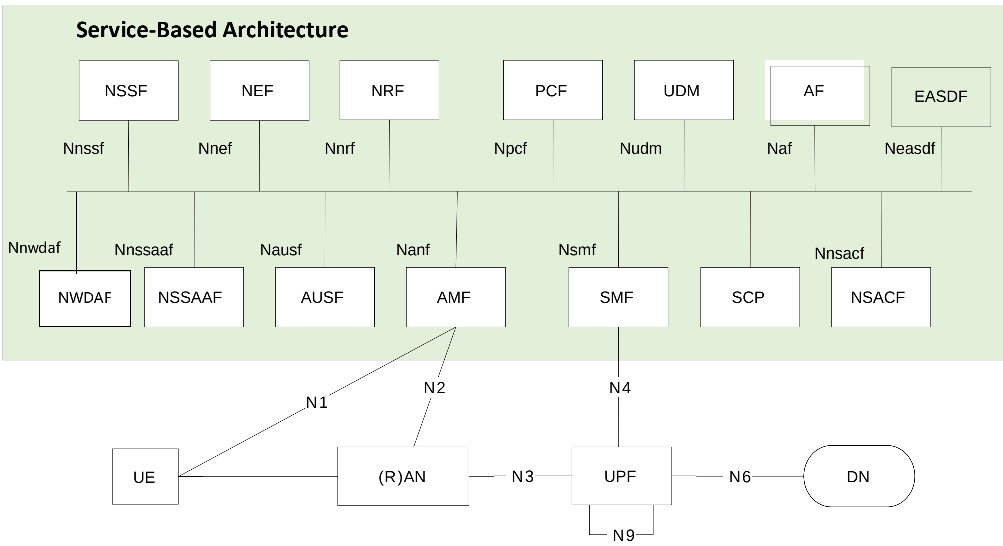

Figure 1. SBA (green box) in a 5G System

The functional features of the SBA in 5G CN include:

- Modular Network Function: the 5G CN is rebuilt into several loosely-coupled NFs, allowing independent scalability, evolution and flexible deployments. NF can provide self-contained, self-managed and reusable services (NF Service, NFS) to other NFs. NF Services are also independently operated and upgraded.
- NF service framework: 5G CN introduces a Network Repository Function (NRF) which is used to provide service registration and service discovery for NFs. As a consumer, a NF uses the NRF to find suitable services. The CN becomes more scalable through this mechanism, without affecting current network when adding and deleting new NFs and NF services.
- Unified Service-based Interface Protocol: The design of the interface protocol in 3GPP considers the combination with information technology and virtualization. Currently, the transport layer of the interface protocol stack is TCP, the application layer is HTTP 2.0, the serialization protocol is JSON, the interface description language uses OpenAPI3.0, and the API design method is RESTful.

The SBA of the 5G CN was first standardized in 3GPP R15[3]. SBA evolved in R16 to eSBA, mainly adding the following functions:

- Service Communication Proxy (SCP). The original direct communication mode between NFS changes to indirect communication mode with the introduction of SCP. Routing functions and service discovery functions can be implemented by SCP through different communication modes. Each NF can pay attention to its own business processing logic, which makes NF functions more lightweight.
- NF/NFS instance sets. A set contains multiple NF/NFS instances of the same type. These NFs or NFSs in a set share context and can access the same data storage. When an NF or NFS fails, services can be seamlessly taken over by other NF or NFSs in the same set, ensuring service continuity and network reliability.

Beyond this foundation, 3GPP R18 studied UPF enhancement for Exposure[4], aiming to support better integration of UPF into the 5GC SBA. Specific features include adding a new SBI for the UPF to open QoS monitoring reporting to PCF, or to open raw state and real-time service flow information to the NWDAF, etc., while the original P2P interfaces with SMF and RAN are still maintained.

1.2 Advantages and Problems

Since SBA for the 5G CN control plane is specified in early releases, it is worthwhile to analyze the benefits and problems of SBA as experienced for CN specifications and deployments.

Advantages

1. Extensibility of NFs and NF Services via NF service framework.

The service-based principles facilitate a dynamic interaction between service producers and consumers: due to service-based design principles, adding a new service of an existing NF or adding a new NF with additional services has a limited impact and dependency to other NFs.

The Network Repository Function (NRF) provides registration and discovery services which allow service producers a (close to real-time) announcement of their new services (or service modifications), while consumers can receive notifications about such modifications immediately. In addition, authorization for service consumption can be managed in a very flexible manner and, if necessary, at a very fine-grained level, e.g., at the level of resources.

2. Cloud-friendly by design.

SBA can adapt to and take advantages of cloud native technologies to allow virtualization and containerization for NFs, creating the benefits of flexible scaling, rapid iteration, improved resource utilization and openness. Service-based architecture intrinsically supports auxiliary services like load balancing and indirect communication, e.g., introducing the SCP functionality that offloads consumers from service discovery and (re-)selection. SCP further minimizes the number of interconnections between NFs that produce and consume services. Besides, 5G Core SBA has also introduced a cloud-friendly protocol stack, e.g., for subscription and notification mechanisms as well as for authentication and authorization purposes.

3. Resiliency and efficiency.

The definition of NF (service) sets provides increased resiliency, e.g., in case of overload or failure of NF instances. Moreover, Unstructured Data Storage Function (UDSF) is used for storing UE context, it allows for flexible extension of data models, enabling product differentiation and hence innovation for optimized performance. Efficiency is further increased by the concept of NF binding, which can significantly reduce the number of service discovery requests.

Problems

1. Performance.

In several deployments, SBA has increased the average latency of the CN control plane procedures, e.g., in cases where a high number of service producers and consumers lead to message sequences that require frequent and repeated service discovery. Moreover, studies seem to indicate that scalability and performance of utilized protocols (e.g., HTTP/2 over TCP) could be enhanced, e.g., see [5].

2. Interoperability.

The high number of supported implementation options, e.g., for indirect and direct communications between service consumers and producers (the standard TS29.500 defines four different models), may lead to interoperability issues.

3. System integration.

The presence of high number of NFs and defined NF services, also considering the continuous addition of new NFs and new NF services, can incur higher integration efforts. Further integration efforts are brought by the possibility to support different implementation options, e.g., direct and indirect communications.

4. Security

Since SBA in CN uses HTTP/2 as the interface protocol, security concerns may be introduced, such as traffic control function security threats caused by improper configuration.

2 SBA extension in RAN

2.1 Related concepts

Historically RAN architecture was mainly designed to guarantee the connection of customers, and the implementation was based on highly integrated software and hardware, with all internal interfaces relatively closed, giving performance advantages of specialization. With the introduction of 5G, mobile networks have enabled more diverse applications and service scenarios and brought the requirements of agility, flexibility and scalability to network operators and users, alike. The transformation of the network to virtualization and cloud-native technologies was an inevitable trend.

The O-RAN ALLIANCE is working to define a cloud-native RAN architecture. RAN functions that 3GPP defined can be deployed in containers through a unified cloud native platform which brings flexibility of management/orchestration capabilities. However, most current protocol functions and processes of the RAN have been specified following the past design philosophy. which is not cloud-native. The implementation of current cloud RAN only changes the running platform for software instead of changing the fundamental software architecture of the RAN. This RAN architecture is not cloud-friendly enough, and cannot make full use of the advantages of cloud-native.

The next step for RAN's evolution towards cloud-native is to define a more adaptable architecture. Although Micro-Service Architecture (MSA) is an important feature of cloud-native, it is more of an application design concept and not very suitable for RAN. RAN can focus on the SBA concepts that are already established in the CN. We use the name “Service-based RAN” to describe RAN architecture designed towards cloud-native principles, having the following features:

- RAN functionalities defined as network services, provided by one or more Network Functions. While their functionality is discretely defined, network services can be bundled together in actual implementations.
- New integration methods for RAN internal services components, and between RAN and CN services, interacting to provide end-to-end functionalities of higher-level services.

2.2 Motivations and requirements

According to the report on framework and overall objectives for IMT-2030 released by ITU[6], diversified requirements and scenarios drive the mobile network future evolution. Service-based RAN is considered a potential direction, with its motivations mainly coming from two aspects: new requirements from differentiated customers and evolving network requirements from network operators. Details are as follows:

- **Experience enhancement for industry customers.** Although 5G introduces support for many new scenarios and use cases and aims to offer better solutions for industries, there are still challenges in real deployments. These challenges include different performance requirements from different industries and customizable, flexible, on-demand service capabilities. It is expected that in future 6G systems, industrial scenarios based on mobile network enablement will be more common, and comparing with the (already service-based) CN, the RAN will play a more important role. A service-based RAN has the advantage of on-demand network function integration, flexible extension, and more agile development.
- **Capability exposure**. With emergence of the Mobile Internet, diversified services need better collaboration between the network and applications. It is desirable for the RAN to expose capabilities to other network functions or to third parties as the current exposure through CN. 6G networks are expected to converge communication, sensing, and AI capabilities. As a result, more demands on network capability exposure will be required. Direct exposure of service-based interfaces, centered around service APIs, can avoid too many P2P interface definitions and unnecessary transparent forwarding among network entities.
- **New service types.** AI, sensing, and other new services will be involved in 6G. The RAN architecture and functional design should integrate new capabilities beyond traditional communication. New service types bring new functions design for both the Control and User Planes (CP/UP), and reduce the complexity for configuration and internal functional interaction. A service-based RAN can easily add/delete services according to the demand, maintaining a uniform design, in any case.
- **Network management.** Operation and maintenance management by telecom operators and other IT support systems are gradually introducing cloud-native design concepts (such as microservices, containers, DevOps, etc.). These concepts meet the rapid development needs of telecom services which includes efficient and reliable platform, agile and rapid operation, and ease of maintenance. As in the case of CNs using SBA in 5G, RANs based upon SBA will help unify network management, reducing the complexity of network operation and maintenance.
- **Network innovation.** Current RAN software updates based on standardized releases have a relatively long life-cycle. This situation cannot meet the needs of various industries that require rapid business iteration. Operators have an increasing demand for rapid innovation of network functions and relatively

lightweight protocol design. Service-based redesign is conducive to the introduction of RAN functions to achieve rapid network iteration. At the same time, its reduced inter-function dependency characteristics help to achieve lightweight design of network protocols and reduce standardization costs.

2.3 Use Cases

The most important features of service-based RAN are agile deployment and on-demand service, and its value is to improve the inclusiveness of the network, which is the biggest difference from the traditional RAN that usually focuses on high performance. Therefore, use cases that require agile deployment, customized capabilities, and rapid iteration are the main application scenarios for service-based RAN.

Industry Customized Networks

The continuous development of industry digitalization requires mobile networks to empower network operators and users, and its scenarios include industrial manufacturing, smart cities, internet of vehicles, and others. Industrial business types vary greatly. Their needs cover every aspect of network performance, deployment, operation and maintenance. The RAN plays a key part in the industrial network deployment. The characteristics of a service-based RAN enable the provision of flexible customized networks for these diverse industries. These characteristics include on-demand RAN functions, dynamic scheduling (of computing, AI, and other resources), agile deployment, and convenient upgrade and maintenance.

Edge Convergence Networks

Based on the development of 5G, it is predicted that in future more businesses with extreme performance demands, such as delay-sensitive classes, will be processed at the network edge (e.g., MEC platform). This will lead parts of the CN and applications to be deployed at the edge. Also evolving computing capabilities may be deployed at the edge for AI and other applications with such requirements, and these new functions will need better integration with RAN. Service-based RAN architecture makes RAN and CN functions convergence at the edge easier through function redesign and will reduce the redundancy. Additionally, for AI or Mobile Edge Computing (MEC) capabilities, SBA is a more flexible way for RAN services to interact with them, and deployment on the same cloud platform is more feasible.

Emergency Communication

In some disasters and emergencies, there is a strong demand for communication recovery and emergency communication. These demands require quick, flexible and adaptable on-demand deployment. Service-based RAN can be rapidly deployed and flexibly expanded in the form of virtualization, while on-demand service composition and dynamic service orchestration can also better ensure the realization of the high-priority communication capabilities.

Sustainable Value-added Networks

The future 6G network is expected to support the user and application trends [6], such as ubiquitous intelligence, integrated sensing and communication, ubiquitous computing, Digital Twin, etc. It can be envisaged that more new value-added features or functions will be introduced into 6G networks. Given that the corresponding technologies may originate in the IT field and often have their own evolutionary routes, 6G network architecture should be openly, flexibly, and sustainably designed in order to be forward-compatible for emerging technologies. From this perspective, a service-based RAN can naturally support the needs of the sustainable development of new value-added network features or functions.

3 Considerations on Service-based RAN

3.1 Principles

Since we take SBA in the 5G CN as a reference, the basic design principles also can follow many aspects of the 5G CN. Some of the most defining principles comprise:

- Modularity of service definition: A reasonably fine-grained definition of services simplifies the introduction of a new services without impacting other services. Hence, modularity also enables service decoupling and reduces dependencies between network functions. Moreover, modularity serves as an enabler for reusing services in various call flows, i.e., invoke them whenever needed.
- Reusability: Producer NFs offer their services via well-defined APIs that can be consumed by any authorized consumer NFs. This not only overcomes redundancy in P2P interface design, but also allows for re-using fundamental services in multiple procedures. Each system procedure can be described by a sequence of service invocations.
- Autonomous and self-contained service: Each of the NF services offered by a Network Function can and should be self-contained, acted upon and managed independently from other NF services offered by the same Network Function (e.g. for scaling, healing, etc.).
- Discoverable services: Consumer NF can discover the services provided by the producer if it is authorized.
- Composable services: When appropriate, different services can be easily combined to provide a composite service that is subsequently accessible through its own SBA APIs.
- Secure access to services: Uniform mechanisms can be used to provide consistently secure access, subject to appropriate access control mechanisms and policies, thus supporting and enforcing secure system design principles such as Zero Trust with minimal impact to the individual NFs and network service implementations.
- Variability: A function which can provide solution differentiation and whose implementations can vary to tune the system performance should be the one which is refactored as a NF service.

Besides, we should also consider the current RAN features to avoid the unreasonable rebuilding of network functions. For example, using the current CU, DU split or CP, UP split as the baseline, RAN procedures should be reused as much as possible and can be treated as service-based entities, allowing for backward compatibility, while not precluding architectural evolution flexibility in the future.

3.2 Overall Framework

In addition to the high-level principles, there are also some other design recommendations for the service-based framework, including:

A common interface framework assures a uniform technical realization of the service interfaces: besides the concrete API design, i.e., protocols and data models, this includes the mechanisms for service consumer authentication and authorization, e.g., using role-based access control mechanisms, subject to operator policies. Further, each NF service can be ensured to be accessible only by means of a well-defined interface (API), where a single interface may consist of one or several operation(s) to allow for more fine-grained or course-grained accessibility of the service capabilities, depending on the service requirements **Error! Reference source not found.**.

Another framework aspect concerns the communication between service producers and service consumers. A “Request-Response” type of interaction allows for synchronous communication, which can be additionally safe-guarded by time out mechanisms. “Subscribe-notify” type of interaction enables asynchronous, event-based communication, e.g., in case when pre-defined thresholds are exceeded or certain failure events occur in the network.

In addition, an abstraction framework like the “Service Communication Proxy” concept defined in 3GPP [7] further separates the NF business logic from the common platform/communication functionality. SCPs support “indirect communication” in the 3GPP CN and take over tasks such as connection management, load balancing, service discovery, and security. An Istio-type sidecar “that mediates inbound and outbound communication to the workload instance it is attached to” [8] comprises a potential realization of the SCP concept.

A service-based architecture also requires focusing on the definition of the specifically defined service interfaces (APIs). The internals of the service-producing NFs can be left to implementation. In this way, individual services can be defined in a self-contained manner, i.e., they operate on their own context and can be handled with their own specific lifecycle management policies. More complex system procedures, such as mobility or registration management, can then be defined by specifying a particular sequence of services to be invoked.

3.3 Service-based RAN functions and interfaces

For the 3GPP 5G RAN architecture, it supports both non-split and split architecture with different interfaces[9]:

- In non-split RAN architecture, a 5G RAN node supports the Xn and NG interfaces.

- In CU-DU split RAN architecture, a 5G RAN CU supports the Xn, NG, and F1 interfaces, and a 5G RAN DU supports the F1 interface.
- In CU-CP and CU-UP split RAN architecture, a 5G RAN CU-CP supports the Xn-C, NG-C, F1-C, and E1 interfaces, and a 5G RAN CU-UP supports Xn-U, NG-U, F1-U and E1 interfaces.

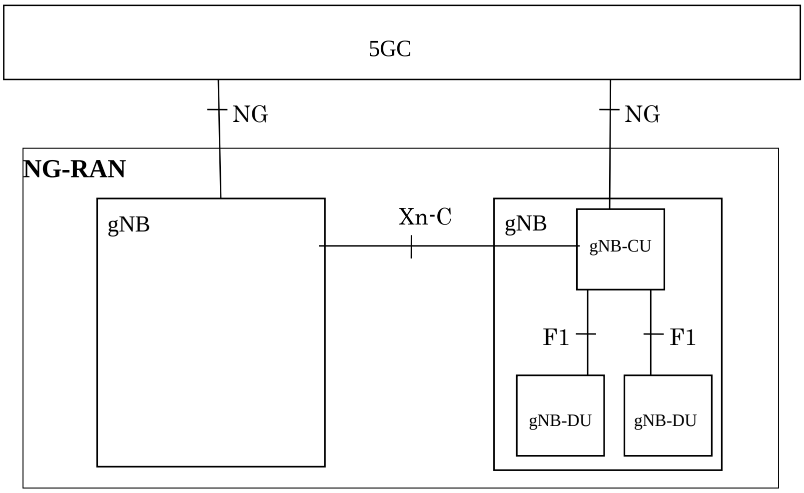

Figure 2. CU DU Split 5G RAN Architecture

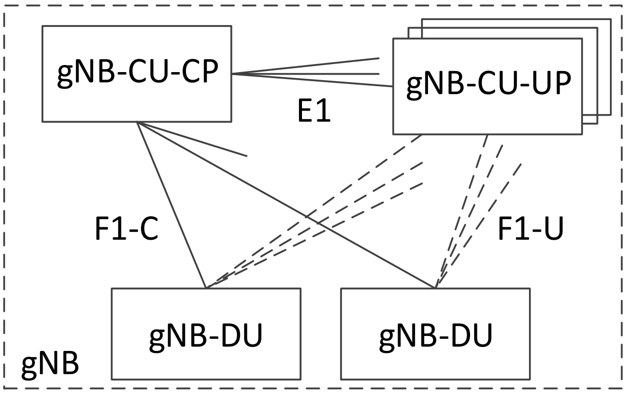

Figure 3. CU-CP and CU-UP Split 5G RAN Architecture

O-RAN defines a more open and disaggregated RAN architecture using the 3GPP split RAN as the baseline, while also introducing more open interfaces for intelligent control and cloud. When defining service-based RAN functions and interfaces, basic 3GPP RAN functions and interfaces can be considered initially, as the service-based architecture is focused on a cloud-native model, and the 6G RAN can also provide AI services, thus requiring these concepts in O-RAN to be included in early stages of work.

For a service-based RAN, the following aspects should be considered:

1. Which conventional 5G RAN functions and interfaces should be rebuilt to be service-based?

2. How best to group the conventional procedures over the conventional 5G RAN interfaces (e.g., Xn, NG, E1, F1), which is also dependent on the required granularity of services?
3. Whether/how to support the future 6G features (e.g., AI, sensing) as new services of the service-based RAN architecture?

Aspect 1: Which conventional 5G RAN functions and interfaces should be rebuilt to be service based?

The 5G RAN CP and UP are designed for different purposes: the CP is used for connection control, radio resource management, etc., while UP is for user data packet processing and transmission. The 5G CU- DU split is based on protocols in which more real-time protocol functions are in DU. As the basis of service-based RAN is providing services for other NFs, it is better to consider rebuilding the functions from the perspective of both the CP and the UP independently. It is also important to evaluate whether the real-time related functions are suitable for implementation using a service-based for the performance concerns, or at least minimally ensuring that performance considerations are taken into consideration in their SB API designs.

For CP service-based redesign, current functions can be rebuilt into finer grained NFs that contain services according to the degree of coupling required. The NFs can realize direct access to the CN through SBI APIs, reducing unnecessary signaling forwarding, and the direct interaction with other CN services can be changed from serial interaction to parallel interaction, optimizing the signaling process of the control plane. The optimization of signaling processes helps improve network performance, such as delay and efficiency.

Besides, different NFs can be combined and flexibly deployed in different scenarios and regions on demand. For example, the mobility and connection related NFs can be deployed in the edge cloud for mobility experiences, but the session management related NFs can be deployed with UP functions near the network edge, depending on differentiated QoS requirements (for example). Future 6G applications will allow CP functions to have more different functional requirements and configurations, for example, the control procedures used for connection or sensing may be different. In this case, service-based redesign can add new services or combine existing services more easily through the granularity of NF services. Finally, for requirements of specific industrial scenarios, it also helps to support RAN and CN integration and deployment at or close to the edge, such as when merging the connection and mobility functions and forming new NFs that simplify deployments, reduce operational complexity, and improve performance.

For the UP, it is technically feasible to refactor the PDCP,RLC, even MAC and PHY layer functions into NF or NFS to achieve service-based and there have been several studies shown these solutions[10][11][12]. However, there are many performance concerns compared to service-based CP, such as packet processing latency, efficiency. This is the reason that SBA in 5G CN is only defined for the CP. Current RAN UP use the defined P2P protocols that comply with the OSI hierarchical protocol design principles. With this design, it is harder to achieve cross-layer function combinations and interactions. 5G uses RRC configuration for different data packet requirements, e.g. PDCP duplication for high-reliability business, PDU set for XR services, etc. For future 6G network services, data packets requirements will vary largely for different businesses that need flexible UP functions, and service-based designs will make different processing order and service combination possible. Considering the requirements and these concerns, it seems more reasonable to support service-based interface over RAN UP as the starting point, then define different UP services combinations based on differentiated business requirements.

Aspect 2: How to refactor the conventional procedures over the conventional 5G RAN interfaces (e.g., Xn, NG, E1, F1), which is also dependent on the required granularity of services?

Many fundamental procedures specified for conventional 5G RAN may be optimized as NFs and can directly interact with each other. They also are expected to be supported in the service-based RAN architecture since they are for basic network functionality, e.g., handover procedures, PDU session modification procedure, etc. However, in a service-based RAN design, they are provided in the form of RAN services that can be invoked by any consumer node such as the AMF o SMF. Grouping those procedures as one service can take different approaches as long as they follow the principles given in section 3.1, i.e., services are autonomous and self-contained, enabling independent deployment, operation, and upgrade.

Aspect 3: Whether/how to support the future 6G features (e.g., AI, sensing) as new services of the service-based RAN architecture?

As described in section 2.3, the 6G RAN is foreseen to support value-added services beyond traditional broadband connectivity, such as AI and sensing. From an architecture design point of view, each value-added feature can be either integrated as a new type of service provided by the same service-based RAN, or can be supported by a dedicated RAN-side network function. For the first approach, current UP and CP functions would need to be enhanced to control the AI or sensing functions and process their new type data flow. Even more functional planes may be introduced in the RAN to complement existing planes, such as data plane, AI, or computing planes. A service-based RAN is easier to update, and it is easier to integrate different NFs/services. For the second approach, new NFs can interact with others through service-based APIs. It is more flexible and elastic for deployment. No matter which approach to take, the tradeoff between flexibility and complexity from implementation point of view should be taken into account.

4 Potential solutions of Service-based RAN architecture

4.1 Solution 1

Based on above considerations, a potential solution for service-based RAN is shown in Figure 4. This solution is backward compatible with the 5G conventional RAN architecture, and the CU-DU split has also been studied in multiple 6G research organizations as a candidate for 6G. Given this solution is arguably easier to implement, it could be the initial stage of the evolution of 6G service-based RAN architecture.

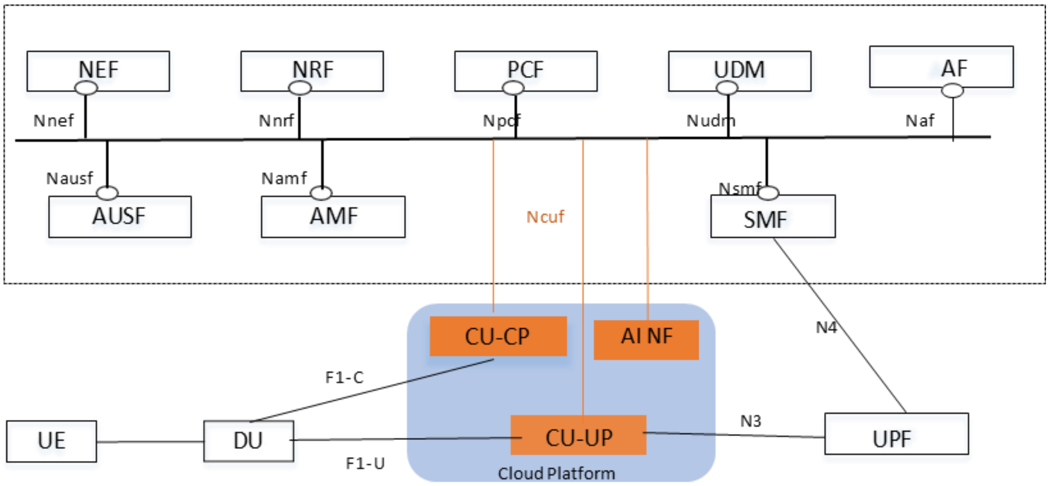

Figure 4. Service-based RAN Architecture

For this solution, the CU-CP and CU-UP are partially service-based and can be treated as NFs. The CU-CP provides the service-based interface to CN NFs and other CU-UPs. However, it also keeps the traditional F1 interface with DU. The CU-UP only implements the service-based E1 to open some capabilities. The P2P interface with UPF and DU are kept intact.

The current CU-CP interacts with other functions through the NG, Xn, and E1 interfaces[13][14][15]. From the perspective of providing services to external network functions, the CU-CP can group these interface functions into services that will contain connection related services, mobility management services, session/bearer management services, NAS message transfer, warning message transfer, location, trace related services, etc. These services need to be exposed by the appropriate Service-Based Interfaces (SBIs). Consumers of these services could include the AMF, SMF, other CU-CP and CU-UP entities, etc. The CU-CP also has some other services needed to provide for the UE operation by F1 and Uu, such as RRC connection and system information, and these functions can be implemented in a service-based manner to support the F1 interface or can be left to implementation. As CU-UP mainly processes and forwards the data packet to the UPF, and the current UPF only considers opening some information through SBI and still keeps the traditional P2P interfaces, so in this design, the CU-UP only considers the E1 interface related functions to be service-based. It will include the bearer related services and trace services. And consumer of this NF could be the CU-CP, SMF, etc.

Additionally, new services such as AI and sensing will be provided by 6G, and the RAN can integrate these functions. This solution introduces these functions as new NFs, deployed with CU in the same cloud, for example, an AI NF containing several services, data collection, model training and inference which can interact with CU and CN AI related NFs through SBIs. As for management and orchestration, there will be a unified framework for the RAN service-based NFs.

The service-based interfaces for the CU-CP and CU-UP may follow the core network SBI protocols that use TCP+HTTP/2+JSON+OpenAPI. There are other research efforts to study other alternative protocols by using HTTP/3+QUIC to deal with network efficiency and congestion problems as Figure 5 shows. SBI will be upgraded as the technology evolves.

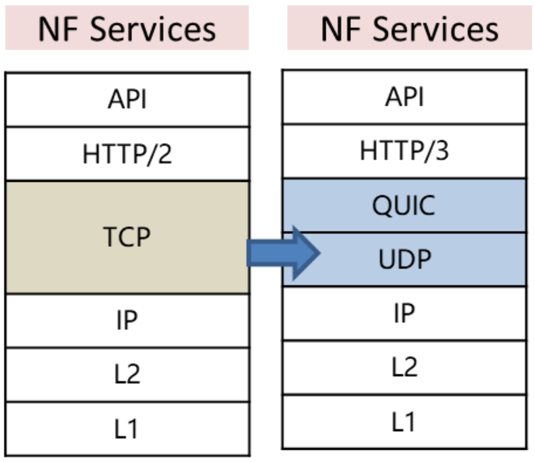

Figure 5. SBI Protocols

For this service-based RAN architecture, RAN NFs such as CU-CP and CU-UP can register to NRF and publish their services to be consumed by other NFs. They can also get other NF services' information. As they can directly invoke the interface to communicate with Core NFs, many procedures can be optimized as the traditional way usually uses the AMF to transfer transparently, e.g., for the session related services, CU-CP can directly interact with SMF. CU-UP can also directly interact with LMF, NWDAF, and NEF for location and data collection.

4.2 Solution 2

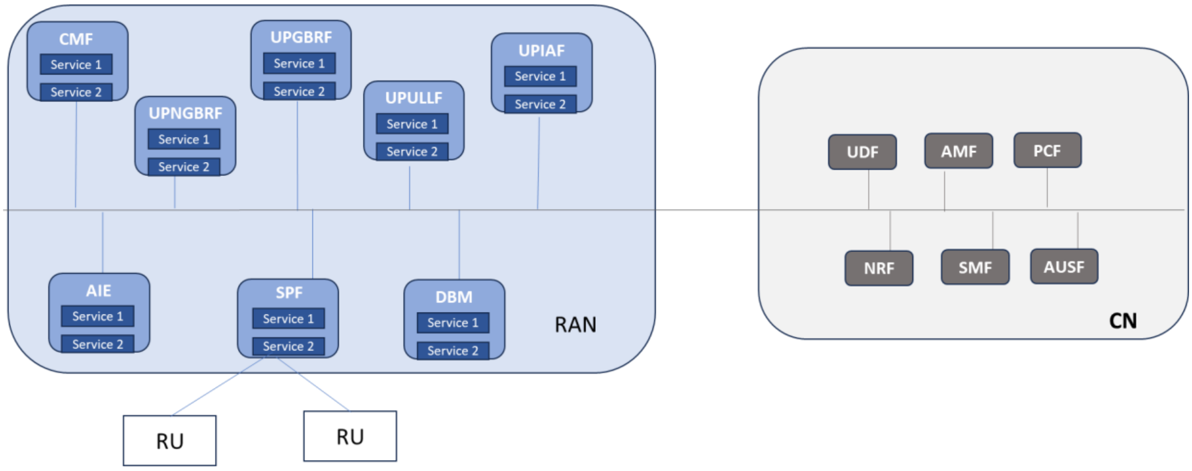

Figure 6. Service-based RAN Solution Based on Functionality

Service-based RAN solution shown in Figure 6 refactors the RAN into NFs based on the end-to-end functionality. Each RAN NF is defined below:

CMF (Connection Management Function): This RAN NF supports the RRC Connection management and session management functionalities such as Paging, system information management and PDU establishment. It exposes services to CN NFs such as AMF/SMF. Also, this NF uses the services provided by UPNGBRF, UPGBRF and UPULLF to configure QoS, PDCP configurations etc.

UPNGBRF (User Plane Non GBR Function): This RAN NF provides the functionality required to support non-GBR traffic. Traditional SDAP, PDCP, RLC protocol functions can be part of this RAN NF. This RAN NF will use the services provided by SPF to transmit/receive data over air interface. UPNGBRF will support the N3 interface functionality towards the UPF if the UPF still follows the current design in 6G.

UPGBRF (User Plane GBR Function): This RAN NF includes functions needed for processing GBR traffic such as voice and video. The corresponding SDAP, PDCP and RLC related processing will be services of this NF. It also will provide services to the UPF for supporting the N3 interface functionality. This RAN NF will use the services provided by SPF to transmit/receive data over air interface. In industrial deployments, there might not be a voice or video requirement and this NF could be removed.

UPULLF (User Plane Ultra Low Latency Function): Functions of User Plane such as NG-U, SDAP, PDCP, and RLC for handling Ultra-low latency traffic will be part of this RAN NF. It will also use SPF services for data transmission/reception over the air interface.

SPF (Scheduler and Physical layer Function): This RAN NF maintains the scheduler and L1 functions. It will provide services to other RAN NFs such as UPNGBRF, UPGBRF, and UPULLF to get the data for transmission at each TTI. Latency requirements are very low for this interaction. SPF will provide services to CMF for transmission/reception of control/paging data.

DMF (Database Management Function): This RAN NF includes the functions of data collection, processing, and distribution, e.g., UE context, QoS information, the cell-level data for load balancing, user-level data for positioning, and AI functions. This NF can be accessed by any other RAN NF.

NCEF (New Capability Enabling Function): This NF is used for RAN new capabilities beyond connection, such as AI, sensing, and computing. For AI, it can support the E2 interface that is compatible with the ORAN architecture. It also can incorporate more AI functions to achieve native AI in RAN in 6G, such as RAN AI control, the interaction with CN and UE AI functionalities, AI modeling training, and interference. It can also be used to control the sensing and computing capabilities.

RCEF (Radio Capability Exposure Function): This NF includes the exposure of some RAN capabilities to 3rd parties, e.g., the real-time performance monitoring data of RAN services, AI capabilities, and computing powers. It is like the NEF in the CN, but the RAN side NF can make some enhancements or tailoring based on RAN side features.

For this solution, different user plane NFs (UPNGBRF, UPGBRF, UPURLLF, etc.) are used for different traffic types. These NFs can be selectively deployed for specific scenarios. Also, NCEF and RCEF can be selectively deployed based on requirements. User Plane NF can also be extended if 6G have more types of QoS features.

Each RAN NF consists of several services to be consumed by others and interactions between different RAN NFs are also based on SBI. By means of dedicated access and security mechanisms, services in the RAN NFs can be selectively exposed to the CN.

4.3 Comparison on different solutions

The biggest difference between the above two solutions lies in the degree of service-based redesign. Solution 1 is more about service-based interface, while solution 2 is a more thorough RAN functional reconstruction.

Solution 1 is more compatible with the existing 5G RAN architecture, conducive to the smooth evolution of 5G, and will not bring too many performance concerns caused by SBA. However, the advantages brought by SBA in this solution are limited, and for new capabilities such as AI, it cannot be better integrated into the RAN.

Solution 2 is a major re-imagining of the existing network architecture, and many processes and functional implementations need to be redesigned, but the advantages brought by SBA, such as flexibility and extensibility, are more obvious. Also, the introduction of new capabilities is easier. However, there will be some challenges caused by SBA.

One option is for solution 1 to be used as the initial attempt of service-based RAN, and then completely evolved to a fully service-based architecture, constantly promoting the implementation of more flexible and customizable RAN, thus enabling more business scenarios for 6G.

5 Challenges

The challenges for Service-based RAN are summarized and explored here. It is expected that these might serve as a basis for future research efforts and refinement on the architecture.

Reasonability of NF/Service definition

Compared with the CN, RAN has more stringent performance requirements, and the structure of RAN functions into NFs/services will inevitably lead to a certain degree of performance loss, despite the best efforts to avoid this issue during system architecture and design. It is important to consider the balance between the performance of existing RAN and the advantages of the service-based features, and to minimize the impacts on performance and efficiency introduced by SBA. Some rules and technologies could be investigated to achieve this goal, e.g., the services can be instantiated/orchestrated to be on adjacent sites rather than the distant ones, how to fully utilize the capabilities of SBA, including such techniques as non-blocking services (request, response), asynchronous services (subscription, notification), etc.

It is undeniable that every split and restructure solution is not perfect but a multi-faceted trade-off such as "effectiveness, inclusiveness, and economy". For future 6G system designs, high performance is not the only optimization goal; agility, customization, adaptation, stability, and other dimensions of network design characteristics, will also become considerations.

Latency

For service-based RAN architectures, SBI needs to support very low latency, e.g., data exchange between SPF and other NFs in solution 2 should be done at a per TTI basis, which could be very challenging. This raises additional requirements for SBI optimization.

Processing

In the service-based RAN architecture, it is important to consider that there will be an extra processing effort as multiple NFs will be involved in handling certain messages. In addition to the execution of the different NF logics, multiple message processing phases will be required, e.g., message (de)serialization. The impact of this extra processing should be considered in terms of extra processing needs at RAN, but also in terms of potential impact to performance for UE and RAN-CN signaling.

Impact on UE

The important point not to be overlooked is the impacts on UE. It seems the less impacts, the better, but the SBA RAN will break the current protocol layers and could affect the UE implementations unless great care is taken (e.g., to avoid UE-visible changes to air interface call flows based on SBA asynchronous logic). The consistency and interactions between UE and Service-based RAN should be analyzed further.

Security

SBA is deemed to be an independent security domain in 5G CN. It is essential to ensure secure communication between NFs. Serviced-based RANs also face similar security challenges, and because the RAN used to be a relatively closed network element, the introduce of new services such as AI and direct capability exposure will bring new security challenges. Furthermore, it is worth investigating the impacts on the existing security mechanism when SBA is introduced in RAN. A further aspect to consider is the implication of SBA in RAN towards the CN, for example, when the exposure of RAN services towards the core network will increase the surface of attacks towards the NFs in the CN.

Interoperability

Different NFs in a Service-based RAN implementation may come from different manufacturers, especially new capabilities such as AI and compute services. Although the interface uses SBI, each manufacturer has its own understanding of the specific implementation of the service, and the workload required for interoperability testing could be significant, even with carefully designed interoperability testing. Interoperability efforts should be considered both among RAN NFs but also between all the RAN NFs which interact with CN NFs. Furthermore, if the UE will be communicating with different RAN NFs, interoperability implications should be analyzed for the UE/air interface side, as now the UE should be tested for interoperability considering different, but valid, implementation models employed for the RAN NFs from different vendors.

6 Conclusion

Service-based RAN is a key element to realize a truly cloud-native RAN. This research report first analyzes the current 5G CN SBA design, including advantages and disadvantages, as a reference for the extension of SBA to the RAN. Next, according to the current design of RAN architecture and 6G requirements, the motivations and use cases for the evolution towards Service-based RAN are analyzed, and some concepts are suggested to be employed in combination with the existing research on cloud RAN. Further, considering the protocol and architecture characteristics of current RAN, the key directions of Service-based RAN are given, and two different solutions are described. Finally, the challenges faced by Service-based RAN are explored.

Service-based RAN is one of the 6G network evolution directions to adapt to the needs of the industry, comply with the trend of technology, and match the goal of network evolution. The report gives a preliminary study, and it is expected that the industry can jointly explore the technology for Service-based RAN, promote the maturity of technology, and create a new eco system for the continuous evolution of the future RAN architecture.

References

[1] 3GPP, TS 23.501 “System architecture for the 5G System (5GS); Stage 2 (Release 18)”, v18.4.0

[2] Snepssneppe M , Namiot D. Micro-service Architecture for Emerging Telecom Applications[J].International Journal of Intelligent Information Technologies, 2014, 2(11):2307-8162.DOI:10.1016/S0140-6736(76)90515-8.

[3] 3GPP, TS23.501, Technical Specification Group Services and System Aspects; System architecture for the 5G System (5GS);Stage 2(Release 15)

[4] 3GPP S2-2202691, Study on UPF enhancement for Exposure And SBA, 2022.

[5] J. Bourbon Moreira et al., “Next generation of microservices for the 5G Service-Based Architecture”, in: International Journal of Network Management, volume 30, issue 6, August 2020

[6] Recommendation ITU-R M. Framework and overall objectives of the future development of IMT for 2030 and beyond.

[7] 3GPP, TS 29.500 “Technical Realization of Service Based Architecture; Stage 3 (Release 18)”, v18.4.0. available at https://www.3gpp.org/ftp/Specs/archive/29_series/29.500/29500-i40.zip

[8] Istio Service Mesh Documentation, “Sidecar”, available at: https://istio.io/latest/docs/reference/config/networking/sidecar/ [retrieved February 16, 2024]

[9] 3GPP TS 38.401 Technical Specification Group Radio Access Network, NG RAN, Architecture description, V18.0.0, (2023-12)

[10] Na Li et al., “Micro-Service-Based Radio Access Network”, China Communications Magazine, March 2022

[11] Zixiao Zhao et al., “Overview of Prospects for Service-Aware Radio Access towards 6G Networks”, Electronics 2022, 11, 1262.

[12] M. Kulkarni, P. Biswas, P. Naik and A. Saha, "A Framework for Design of Portable and Scalable Physical Layer in O- RAN," 2024 16th International Conference on Communication Systems & NETworkS (COMSNETS), Bengaluru, India, 2024

[13] 3GPP TS 38.413. “Technical Specification Group Radio Access Network; NG-RAN; NG Application Protocol (NGAP)”, V18.1.0

[14] 3GPP TS 38.423. “Technical Specification Group Radio Access Network; NG-RAN; Xn application protocol (XnAP)”, V18.1.0 (2024-03)

[15] 3GPP TS 37.473.“Technical Specification Group Radio Access Network;E1 Application Protocol (E1AP)” ,V17.8.0, (2024-03)