O-RAN next Generation Research Group (nGRG)

Contributed Research Report

Research Report on RAN-CN Converged Architecture

Report ID: 2024-13

Contributors:

CTC

Lenovo

Nokia

Release date: 2024.02

Authors

Jiaying Zong, CTC (Editor-in-Chief)

Shuigen Yang, Lenovo

Hannu Flinck, Nokia

Reviewers

Douglas Knisely, Qualcomm Technologies, Inc

Tao Chen, VTT Finland

Shan Liu, China Unicom

Sandeep Sharma, Tech Mahindra

Ravi Sinha, Reliance Jio

Disclaimer

Copyright

Executive summary

6G networks aim to further enhance wireless communication performance, expand capabilities such as Artificial Intelligence (AI) enabled Continuous Integration and Continuous Deployment (CI/CD) automation and perception, and broaden the scope of network applications, providing more advanced and comprehensive support for the future digital society. In the process of network design, it is necessary to consider more efficient ways to meet different application requirements. Especially for industry applications which have requirements of data/network control signalling staying within a specific area or lightweight network deployment, certain functions of the Core Network (CN) (such as user plane functions and control plane functions) will benefit from being decentralized to the edge site for highly distributed deployment in order to ensure specific metrics such as performance and security. Additionally, integrating some elements of the CN with the Radio Access Network (RAN) can improve system performance and resource utilization while reducing network complexity and costs by minimizing unnecessary forwarding and processing.

This report explores the RAN-CN converged architecture as we move towards 6G. It begins with a brief introduction to the current RAN and CN architectures involved in the 3rd Generation Partnership Project (3GPP) and O-RAN. The report then delves into the motivations and requirements for RAN-CN convergence, highlighting the shortcomings of the current architecture and the potential needs for future network evolution. Additionally, it provides a clear definition of RAN-CN convergence and analyses various use cases, such as enterprise private networks, edge mobility, edge intelligence, edge massive access, resource sharing and cooperation, and shared data analytics and exposure.

The RAN-CN converged reference architecture is derived by first studying the solution constraints. Additionally, the design principles are analysed, focusing on principles of network simplification and reduction of overheads, flexibility, configurability, scalability, compatibility, security, and privacy. Based on these considerations, the potential converged functionalities between RAN and CN are depicted. Specifically, session management, state management, mobility management functionalities are considered for the potential convergence. Subsequently, a potential RAN-CN converged architecture is proposed based on the End-to-End (E2E) Service-Based Architecture (SBA). Furthermore, the impacts of RAN-CN convergence on the existing O-RAN and 3GPP framework are analysed in terms of architecture, interfaces, protocol stacks, and signalling procedures. Finally, challenges related to RAN-CN convergence such as security, privacy concerns, and standardization within the industry ecosystem are investigated.

The convergence of RAN-CN is a promising direction for the future development of 6G networks. This report provides preliminary study on RAN-CN convergence and serves as a reference for 6G network architecture design.

Table of Contents

List of abbreviations

3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
AMF	Access and Mobility Management Function
AR	Augmented Reality
AS	Access Stratum
AUSF	Authentication Server Function
BS	Base Station
CAPEX	Capital Expenditure
CI/CD	Continuous Integration and Continuous Deployment
CM	Connection Management
CN	Core Network
DN	Data Network
E2E	End-to-End
eMBB	Enhanced Mobile Broadband
FCAPS	Fault, Configuration, Accounting, Performance, Security
GWCN	Gateway Core Network
IIoT	Industrial Internet of Things
MAC	Medium Access Control
MDAS	Management Data Analytics Service
mMTC	Massive Machine Type Communication
MOCN	Multi-Operator Core Network
NAS	Non-Access Stratum (NAS)
NEF	Network Exposure Function
NF	Network Function
Near-RT	Near-Real-Time
Non-RT	Non-Real-Time
NSSF	Network Slice Selection Function
NWDAF	Network Data Analytics Function
O-CU-CP	O-RAN Central Unit Control Plane
O-CU-UP	O-RAN Central Unit User Plane
O-DU	O-RAN Distributed Unit

OPEX	Operating Expenditure
O-RU	O-RAN Radio Unit
PCF	Policy Control Function
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
PHY	Physical
PNF	Physical Network Function
QoE	Quality of Experience
QoS	Quality of Service
RAI	RAN Analytics Information
RAN	Radio Access Network
RIC	RAN Intelligent Controller
RLC	Radio Link Control
RNA	RAN Notification Area
RNC	Radio Network Controller
RRC	Radio Resource Control
SBA	Service-Based Architecture
SBI	Service-Based Interface
SDAP	Service Data Adaption Protocol
SMF	Session Management Function
SMO	Service Management and Orchestration
ToC	To Consumer
UE	User Equipment
UDM	Unified Data Management
UPF	User Plane Function
uRLLC	Ultra-Reliable Low Latency Communication
V2X	Vehicle-to-everything
VNF	Virtualized Network Function

List of figures

1 Background

1.1 Current RAN and CN architecture overview

The mobile telecommunications system provides voice, data, and multimedia services to users. Its architecture consists of four main components: the CN, the transport network, the RAN, and the User Equipment (UE). The CN achieves the switch of voice calls, texts, data, and more. Meanwhile, the RAN handles the radio communications between the network and the UE. The transport network connects the CN and the RAN.

1.1.1 RAN and CN Architecture in 3GPP

The 3G communication system adopted three-level architecture including the Node B, the Radio Network Controller (RNC) and the CN. The Circuit Switched (CS) domain for voice and the Packet Switched (PS) domain for data are both supported by the CN.

To decrease the E2E delay, two-level architecture, containing the eNodeB and the CN was proposed in 4G communication system. The functions of the RNC were partly divided between the eNodeB and the CN. The 4G CN evolved from separate CS and PS domains to one common IP domain.

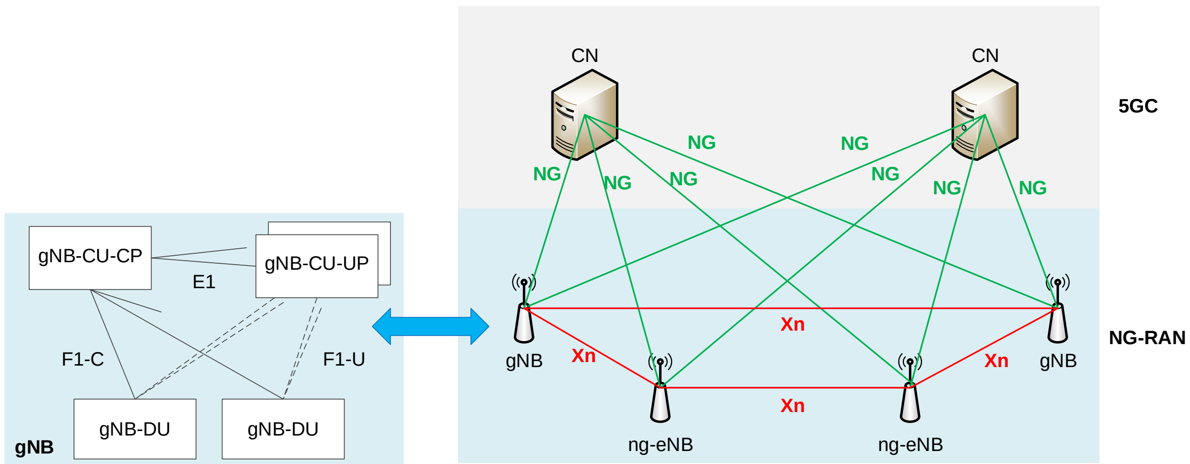


Figure 1-1: Overall Architecture [1] [2]

Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC), and Ultra-Reliable Low Latency Communication (uRLLC) represent the three fundamental use case categories of 5G. These categories define the core requirements that 5G must meet and drive innovations in its architecture. The overall architecture of the 5G RAN as defined by 3GPP is depicted in the Figure 1-1. The functions of the gNB/ng-eNB encompass radio resource management, data compression and decompression, encryption and decryption, integrity protection, user plane data routing, control plane information routing, connection setup and release, scheduling and transmission of paging and system broadcast information, session management, and more. The gNB can be further divided into CU (Centralized Unit) and DU (Distributed Unit), which helps to reduce deployment costs and enhance network flexibility. The CU can be split into CU-Control Plane (CP) and CU-User Plane

(UP) to increase the flexibility of network management. A gNB can consist of one gNB-CU-CP, multiple gNB-CU-UPs, and multiple gNB-DUs.

The Figure 1-2 below illustrates the architecture of a non-roaming 5G system. The CN is divided into control plane and user plane components. The control plane includes network functions such as AMF, Session Management Function (SMF), Authentication Server Function (AUSF), Network Exposure Function (NEF), Policy Control Function (PCF), Unified Data Management (UDM), etc. The CN employs the SBA, where the control plane functionalities and common data repositories in a 5G network are provided through a set of interconnected Network Functions (NFs). Each NF is authorized to access and utilize the services of other NFs. The connection between RAN and CN control plane is established solely through the N2 interface. As a result, interactions between other CN functions and RAN need to be forwarded through the AMF. Additionally, based on the N2 interface, the UE can establish a logical N1 interface with the AMF to facilitate NAS signalling interactions.

The user plane element is the UPF, which is responsible for serving as the mobility anchor, serving external Protocol Data Unit (PDU) session points connected to the data network, data routing and forwarding, user plane Quality of Service (QoS) processing, traffic usage reporting, and more. The UPF communicates with the SMF in control plane through a point-to-point N4 interface. Simultaneously, the UPF connects to the Data Network (DN) through the N6 interface to access external network services. The connection between RAN and CN user plane is established through the N3 interface.

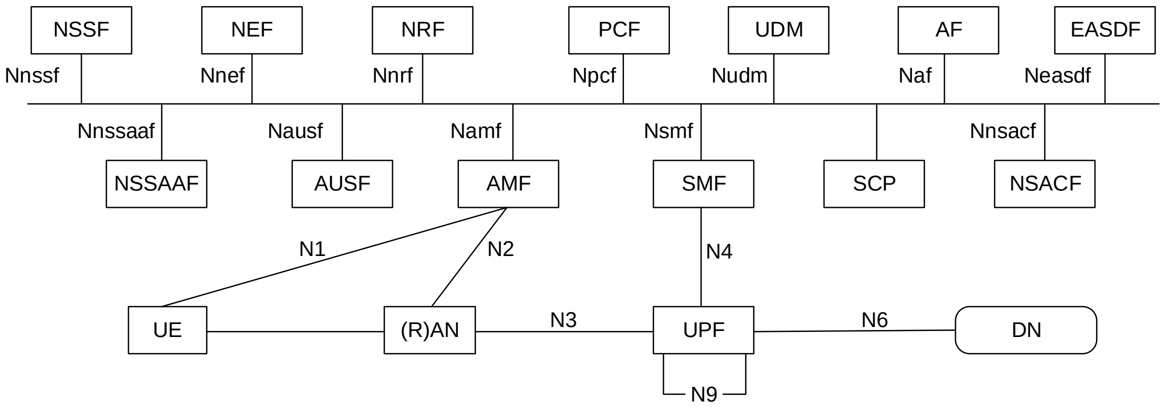


Figure 1-2: Non-Roaming 5G System Architecture [3]

1.1.2 RAN Architecture in O-RAN

Based on the 5G RAN defined by 3GPP, O-RAN introduces an open concept to promote interoperability of hardware and software from different vendors, enhancing flexibility and innovation. The logical architecture of O-RAN is illustrated in Figure 1-3. O-RAN network functions can be virtualized as Virtualized Network Functions (VNFs), represented by virtual machines or containers, positioned above O-Cloud and/or Physical Network Functions (PNFs) utilizing custom hardware.

Currently, Service Management and Orchestration (SMO) is primarily responsible for managing the RAN domain. It provides Fault, Configuration, Accounting, Performance, and Security (FCAPS) interfaces to O-RAN functions, conducts Radio Access Network

(RAN) optimization through Non-Real-Time RAN Intelligent Controller (Non-RT RIC), manages RAN cloud infrastructure (O-Cloud), orchestrates workflows, and handles RAN-related tasks.

On the radio side, components include the Near-Real-Time (Near-RT) RIC, O-RAN Central Unit Control Plane (O-CU-CP), O-RAN Central Unit User Plane (O-CU-UP), O-RAN Distributed Unit (O-DU), and O-RAN Radio Unit (O-RU) functions. The O-DU and O-RU are connected via an open fronthaul interface. The Near-RT RIC interacts with RAN functions through the E2 interface, collaborates with the Non-RT RIC in the SMO via the A1 interface, and provides RAN Analytics Information (RAI) to Y1 consumers through the Y1 interface.

O-Cloud is a cloud computing platform that aggregates physical infrastructure nodes to cater to the demands of O-RAN functions. The cloud platform hosts relevant O-RAN components such as Near-RT RIC, O-CU-CP, O-CU-UP, and O-DU. O-Cloud supports software components including operating systems, virtual machine monitors, container runtimes, and suitable management and orchestration functionalities.

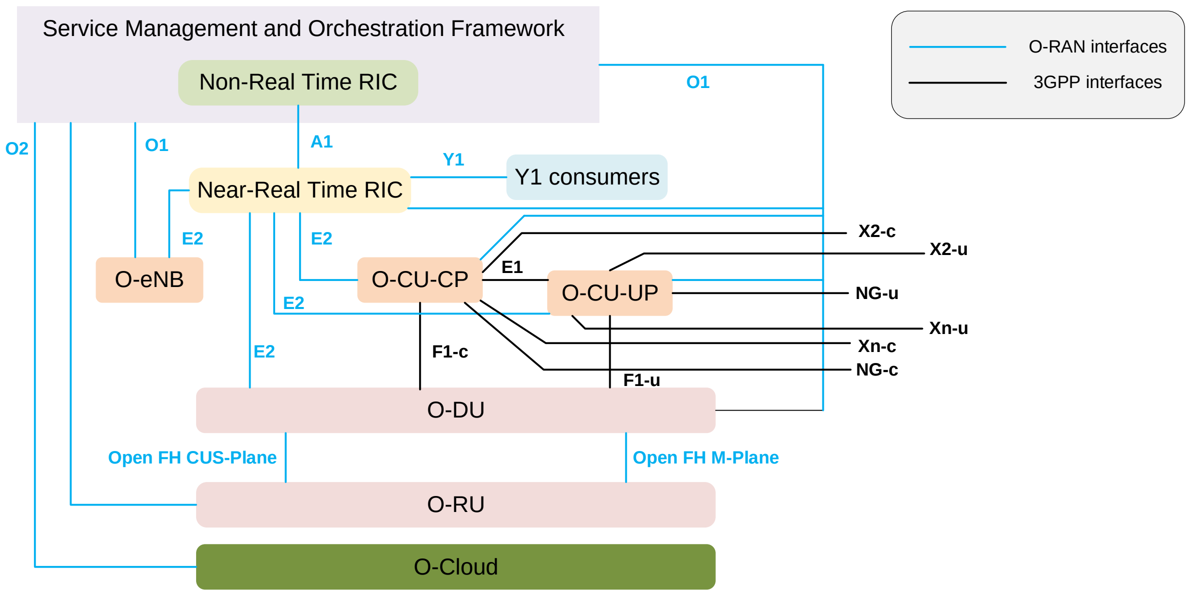


Figure 1-3: Logical Architecture of O-RAN Error! Reference source not found.

1.2 Motivation and requirements for RAN-CN convergence

1.2.1 Description

RAN-CN convergence refers to the seamless integration and coordination between the RAN and CN functions, which represent a design paradigm where these traditionally separate entities work in a tightly coupled manner to optimize network performance, flexibility, and resource utilization. The convergence can occur at the deployment or functionality level:

- **Deployment convergence**: This includes the physical co-deployment of various RAN and CN functions, where the functionalities, interfaces, and internal processing logic of RAN and CN remain relatively independent.

- **Functionality convergence**: This includes the integration and convergence of logical functions and corresponding resources between RAN and CN. The objective is to reduce signalling and processing overhead, improve network performance, optimize resource utilization, cater to diverse business requirements, and provide better support for emerging technologies and applications.

1.2.2 Requirements

The demands for RAN-CN convergence arise from the limits form the current architecture and the new requirements for future network architecture evolution. The existing architecture has the following issues.

- **Limited support for virtualization and cloud-native features.** The current RAN is structured with a hierarchical point-to-point protocol stack. There is a clear boundary between the RAN and the CN. They interact through point-to-point N2 and N3 interfaces for control and user plane exchanges, and these interfaces are not inherently cloud-native as they were designed well before cloud-native technologies existed. The current architecture and interaction mode cannot fully leverage the advantages of network virtualization and cloud-native technologies.
- **Functionality duplication, similarity, or redundancy.** There are many similar or redundant functions in both RAN and CN, such as mobility management, session management, reachability management, paging, state management, and so on. These similar functions require frequent interaction and coordination between the RAN and the CN, leading to complexity and overhead that are not conducive to future network evolution.
- **High inter-NF dependencies among network functions**. The dependencies among CN functions and between CN functions and RAN nodes can result in lengthy and complex signalling procedures, increasing the likelihood of error conditions and introducing unnecessary complexity. This can weaken the system's scalability for future advancements [5] .
- **Efficiency issues caused by single point forwarding**. Control information exchanged between UE and various CN elements, e.g. between the UE and the SMF, needs to pass through both the CU-CP and AMF anchor points, leading to increased latency, computational overhead, and signalling costs.

Moreover, RAN-CN convergence can better support future network evolution.

- **Performance Improvement**. Enhancing performance indicators such as ultra-high reliability and ultra-low latency requires the collaboration or convergence of RAN and CN. For example, in deterministic networks, cross-domain collaboration and scheduling may be an attractive way to ensure E2E deterministic services.
- **RAN Functionality Evolution and Expansion towards 6G**. By integrating with the CN, the RAN can facilitate more flexible functional extensions for 6G,

e.g., supporting the development of new and innovative network services. This may include RAN openness through the NEF or its enhanced functionalities, as well as RAN network intelligence through Network Data Analytics Function (NWDAF) or its enhanced functionalities.

2 Use Cases of RAN-CN Convergence

The convergence of RAN and CN can reduce unnecessary interactions between the current RAN and CN, enabling more efficient resource coordination. This leads to reduced latency, enhanced operational efficiency, and optimized resource utilization. This section presents several use cases that demonstrate the need for RAN-CN convergence.

2.1 Private Networks

To meet the specific needs of industries such as energy, manufacturing, and transportation, private networks have emerged. These networks are designed to provide differentiated and optimal network performance, network and data security, and more. Different private networks have specific requirements in terms of network performance, including reliability, throughput, latency, and jitter.

In terms of network deployment, private networks often support the deployment of lightweight CNs to ensure the security of private network data and privacy (for example, by physically incorporating the CN within and edge infrastructure on the customer's premises under complete control of the private network operator). At the network function level, the co-deployment of the CN and the RAN further enables the integration and reconstruction of logical functions between the two, supporting more efficient and streamlined network capabilities while optimizing network performance (for instance, session management functionalities can be converged between the RAN and the CN to enable fast and efficient session control). The functional convergence allows for the ability to support the evolution of new services and applications.

2.2 Seamless Mobility and Low-Latency Support at the Edge

In the following scenarios, it is necessary for the network to better support the mobility of terminals or users at the edge. RAN-CN convergence at the edge, e.g. converging the mobility management functionalities, can boost performance, reduce latency, and ensure a more efficient, responsive network for edge applications.

- Autonomous Driving, Vehicle-to-everything (V2X) and Intelligent Transportation: Autonomous vehicles require real-time reception and processing of a large amount of sensor data, as well as fast decision-making and adjustments while in motion. In the context of V2X communication, vehicles need to stay connected to the network for real-time navigation, traffic information retrieval, vehicle monitoring, and other functionalities. The edge network needs to support seamless handovers between different BSs to ensure continuous connectivity and high-quality communication.
- Industrial Internet of Things (IIoT): In IIoT scenarios, devices are often distributed in environments such as factories, warehouses, or production sites.

These devices may move or be redeployed frequently, and the network needs to support seamless handovers between different access points to maintain continuous communication and real-time data transfer.

- Aircrafts: The network needs to support the high-speed mobility and communication requirements of unmanned aerial vehicles, airplanes, and other aircrafts, enabling real-time processing and data analysis during flight.
- Mobile Gaming and Augmented Reality (AR): Mobile gaming and AR applications require low-latency interaction and high-quality image rendering. The edge network can support device mobility to provide a better gaming experience and AR effects.
- Mobile Healthcare: In mobile healthcare scenarios, medical devices and sensors often need to communicate with the edge network to transmit patients' health data and monitoring information. As patients may move within a hospital, home, or other mobile environment, the network needs to support fast device switching and seamless transfer of connections to enable real-time monitoring and remote healthcare services.

The above scenarios highlight the importance of seamless mobility and real-time processing support at the edge network, which plays a crucial role in improving device availability, system stability, and user experience. The convergence or reconstruction of mobility management functions in RAN and CN can be considered to simplify functionality and signalling transmission, achieving more concise, efficient, continuous, and high-quality service provision.

2.3 Edge Intelligence

The fields of intelligent transportation, smart campuses, smart manufacturing, AR/VR, and smart homes require extensive data transmission, stringent security, very low latency, and privacy protection measures, as well as real-time processing and analysis to provide intelligent decision-making. The network should therefore deploy intelligent functionalities at the edge to avoid excessive data transmission and prevent overloading the central network, enhance security, and reduce response times. Given that both RAN and CN require intelligent capabilities, it is possible to consider converging or reconstructing the intelligent functionalities of RAN and CN at the edge. This can further simplify the network, reduce latency, improve application response speed and enhance user experience.

2.4 Massive Access and Processing Support at the Edge

Industrial control, V2X, smart healthcare, and smart living are IoT scenarios characterized by large-scale device connectivity and the need for low-latency operations. These require the network to provide the capability for massive device connectivity and processing at the edge. In addition, to meet the requirements of ultra-low power consumption, extremely small size, and low cost, zero-power communication is being considered to support industrial sensor networks, smart warehousing, wearable devices, smart homes, and other applications [6] . These scenarios pose demands for simplification of the network architecture at the edge. Firstly, there is a need for simplification of logical functionalities, such as considering the convergence of RAN and CN access control functionalities. Secondly, there is a need for simplification of signalling interactions. The convergence of logical functionalities facilitates subsequent simplification of signalling processes. These approaches also enable data sovereignty, making it a significant differentiator for enterprise private networks.

2.5 Resource Sharing and Cooperation

The following use case scenarios involve resource sharing and collaboration between the RAN and the CN to provide enhanced performance. The resources can include computing resources, storage resources, processing capabilities, etc. The convergence or reconstruction of RAN and CN functionalities enables the joint optimization.

- Inter-RAT switching: In environments with multiple wireless technologies coexisting, resource sharing and collaboration between RAN and CN are needed to achieve seamless network switching and ensure uninterrupted connectivity for users across different technology networks.
- Multi-layer network deployment: In complex network deployment scenarios, such as indoor & outdoor coverage, high & low frequency coverage and narrowband & broadband radio access, resource sharing and collaboration between RAN and CN are necessary to optimize network coverage and capacity, providing a better user experience.
- Special application scenarios: Certain special application scenarios, such as public safety and medical emergency response, have high requirements for network reliability and real-time capabilities. Through resource sharing and collaboration between RAN and CN, more reliable network connections and timely data transmission can be provided to meet the demands of these special application scenarios.

2.6 Shared Data-Analytics and Exposure

Network Data Analytics and AI/ML are expected to play a significant role in achieving cost efficient operations and optimal E2E performance in 6G. Currently, in 5G networks data analytics functions are optional with limited cross-domain/plane support. In fact, network management is providing one to one communication with Core Network analytics and RAN (i.e., in management plane management and data analytics (MDAS, [7]), in Core Network NWDAF [8] and in RAN RAN-DAF). In order to provide efficient closed control loops to cover E2E performance, RAN and CN analytics should be able to interact directly as well either over a common cross-domain data sharing and exposure framework or by means of using service exposure APIs of the respective data analytics producer in a cross-domain manner. RAN-CN convergence enables direct interaction and collaboration between the analytics functionalities of the RAN and CN.

3 Potential solutions for RAN-CN converged architecture

3.1 Design Constraints

When designing RAN-CN convergence, certain constraints can be considered based on the critical characteristics of 5G RAN. In this section, we use network sharing and network slicing as examples to analyze the design constraints of RAN-CN convergence.

3.1.1 Network Sharing

Network sharing can include the following categories defined in 3GPP [9] [10] .

- Passive RAN sharing, also known as infrastructure sharing (including site sharing).
- Active RAN sharing, where active network elements of the RAN are shared:
  - RAN-only sharing (MOCN), i.e. eNodeBs in an E-UTRA network, and gNBs in a 5G NR network;
  - Gateway Core Network (GWCN), in which not only the RAN elements are shared but also part or all of the CN elements.

For E-UTRAN both MOCN configuration and GWCN configuration are supported over the S1 reference point [11] . In Release 18, only the 5G MOCN network sharing architecture, in which only the RAN is shared in 5G System, is supported [3] .

The consideration of network sharing remains crucial for RAN-CN convergence towards the next-generation network. This aims to maximize network deployment efficiency while minimizing the Capital Expenditure (CAPEX) and Operating Expenditure (OPEX), ultimately enhancing overall network quality. The network sharing approach in RAN-CN convergence, as depicted in Figure 3-1, requires careful consideration, particularly because 5G currently does not support GWCN. The design of RAN-CN converged architecture should take into account network sharing, which involves determining which converged functionalities can be shared, how the shared converged functionalities connect to different operators' CNs via which interfaces.

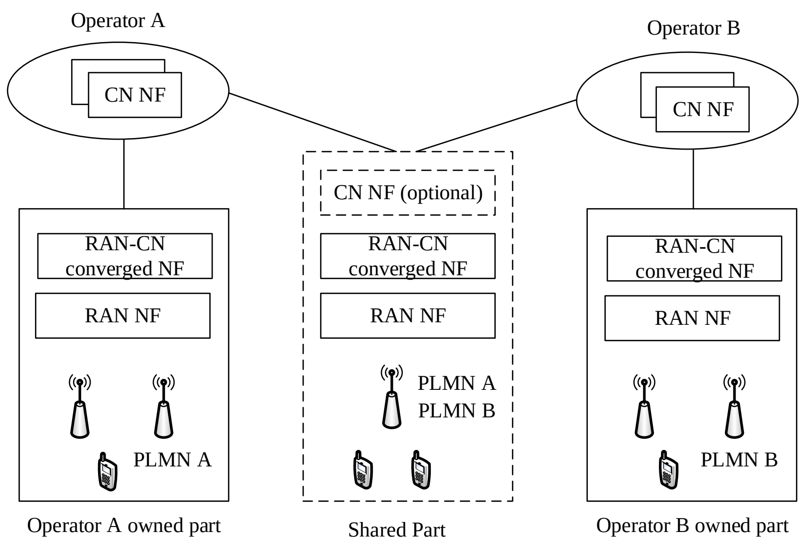


Figure 3-1: Network Sharing Reference Architecture for RAN-CN convergence

3.1.2 Network Slicing

Network slicing refers to dividing a physical network into multiple virtual networks logically, aiming to provide customized and differentiated services for specific applications and subscribers. 5G network slicing comprises both RAN part and CN part.

5G RAN can realise the different network slices by scheduling and also by providing different L1/L2 configurations. As defined in 3GPP, several key principles [1] apply for support of network slicing in NG-RAN, including RAN awareness of slices, selection of RAN part of the network slice, RAN selection of CN entity, resource management between slices, resource isolation between slices, and so on.

For 5G CN, network slicing can encompass both control plane and user plane functions. Certain CN functions (such as AMF, UDM, Network Slice Selection Function (NSSF)) can be shared across multiple slices, while others (e.g. SMF, UPF) can be deployed and customized based on the requirements of different slices.

How the RAN-CN converged architecture supports network slicing and what enhancements and improvements are needed should be considered. For instance, whether integrating similar functionalities of RAN and CN in the converged network may affect the flexibility of network slicing. Additionally, during the design process of the RAN-CN converged architecture, consideration should also be given to the impact on network slicing architecture.

3.2 Design Principles

The design of the 6G RAN-CN converged architecture can consider the following potential principles.

1) **Network Simplification and Reduction of Overheads**
   The RAN-CN convergence should stem from objectives such as network simplification, overhead reduction, efficiency enhancement, cost and energy reduction, and so on. Meanwhile, the impact and workload that the RAN-CN convergence may introduce to the current architecture should be considered and analyzed.
2) **Flexibility and Configurability**
   The architecture should support flexible network configurations and functional deployments, enabling operators to customize the network according to their demands. This may include dynamic resource allocation and swift deployment of new functionalities.
3) **Scalability and Compatibility**
   The RAN-CN converged architecture should adapt to diverse scenarios, dynamically accommodating the continuous evolution of future networks. Furthermore, it can be compatible with non-RAN-CN converged architectures and enable seamless interaction.
4) **Security and Privacy**

The converged architecture design can consider relevant security and privacy mechanisms, such as whether certain CN data can be accessed by the RAN side. This ensures the privacy and security of user and network data.

5) **Data Sovereignty**

Data sovereignty can be a key consideration when designing the RAN-CN converged architecture, as cybersecurity, data protection, and privacy are critical to enterprise operations.

3.3 Analysis of Functional Convergence Points Between RAN and CN

Based on the above analysis, we consider which functionalities between RAN and CN have the highest potential and would receive the most benefits from convergence in the 6G architecture.

1) **Session management functionality**

For the current status of 3GPP, both RAN and 5GC have PDU session management functionality, which leads to some functional redundancy. For example, in the process of establishing a PDU session, the RAN side needs to allocate resources to support the PDU session and corresponding QoS flows between the UE, RAN, and CN. The SMF in 5GC is responsible for session establishment, including maintaining tunnels between the UPF and RAN. Due to the independent deployment of session management functionality in RAN and 5GC, the base station needs to interact with UE and SMF/UPF separately to establish the PDU session, which can result in low efficiency and high complexity. Furthermore, in the 5GC, the PDU Session Management functionality is handled by the SMF. However, the signalling interactions between the UE and the SMF require forwarding through both the RAN and the AMF, resulting in additional latency and forwarding overhead. For example, the PDU Session Establishment Request message sent by the UE to the SMF needs to be forwarded through the RAN and AMF.

Therefore, we can consider exploring the restructuring and convergence of the session management functionality corresponding to the CN and RAN.

2) **State management functionality**

In the current 5G network defined by 3GPP, the connection management (CM) primarily refers to management of the establishment and release of NAS signalling connections between UE and AMF over the N1 interface, including the signalling connections between UE and RAN, as well as between RAN and AMF over the N2 interface. The CM state includes CM-IDLE and CM-CONNECTED. The Radio Resource Control (RRC) state management between UE and RAN includes RRC-IDLE, RRC-CONNECTED, and RRC-INACTIVE.

There is a coupling relationship between the CM state and the RRC state. For example, if the RRC state is Idle, indicating that the connection between UE and RAN is not established, it cannot enter the CM-CONNECTED state. Therefore, UE can exist in various states, such as RRC_IDLE/CM-IDLE, RRC_CONNECTED/CM-IDLE, RRC_CONNECTED/CM-CONNECTED, RRC_INACTIVE/CM-CONNECTED, etc. The complexity and overhead of state management are high, and there is a tight coupling between RAN and CN states. Therefore, the restructuring and convergence of the state management functionality corresponding to the CN and RAN can be explored.

3) Mobility management

(i) Handover

Handover between NG-RAN nodes based on Xn interface and N2 interface requires cooperation between RAN and CN in the current 3GPP specifications. The involved CN functions include AMF, SMF, UPF, etc. The AMF not only provides handover-related functionalities but also acts as a relay for RAN-CN control plane data forwarding, which makes the process complex. Therefore, in the future, RAN-CN handover functionality may be considered for convergence or reconstruction based on specific requirements.

(ii) Tracking area update and paging

In the current 3GPP specifications, to support UE mobility, the CN has the capability for mobility registration update, while the RAN has the capability for RAN Notification Area (RNA) update. For UE in CM-CONNECTED and CM-IDLE states, when it moves to a new tracking area outside its registered area, a mobility registration update is triggered **Error! Reference source not found.**. For UE in RRC_INACTIVE state, it remains connected to the CN and in CM-CONNECTED state. When the cell reselection process selects a cell that does not belong to the configured RNA, an RNA update is triggered.

Correspondingly, for UE in RRC_IDLE and RRC_INACTIVE states, the network may perform paging in the registered area/tracking area/RAN tracking area. Paging for UE in RRC_IDLE state is initiated by the CN, while paging for UE in RRC_INACTIVE state is initiated by the RAN [1] .

Based on the above analysis, there is a certain level of functional redundancy in RAN and CN, and these functionalities are related to UE states. In order to meet specific future application requirements, such as network simplification and enhanced UE mobility, RAN-CN functionality convergence should be considered.

3.4 Potential RAN-CN converged architecture

A potential RAN-CN converged architecture is the SBA, which enables more flexible, scalable, and efficient deployment and management of network functions and services. With SBA, network functions can be developed and deployed independently, and they can easily connect and form E2E service chains.

Since SBA was already introduced into the CN, the introduction of SBA into the RAN will bring an E2E SBA, where RAN can communicate directly with other NFs via service-based interface. Figure 3-2 shows an example of the E2E SBA. The E2E SBA refers to the application of SBA principles throughout both the CN and the RAN, ensuring the NFs in the RAN and the CN cooperate in a service-based manner.

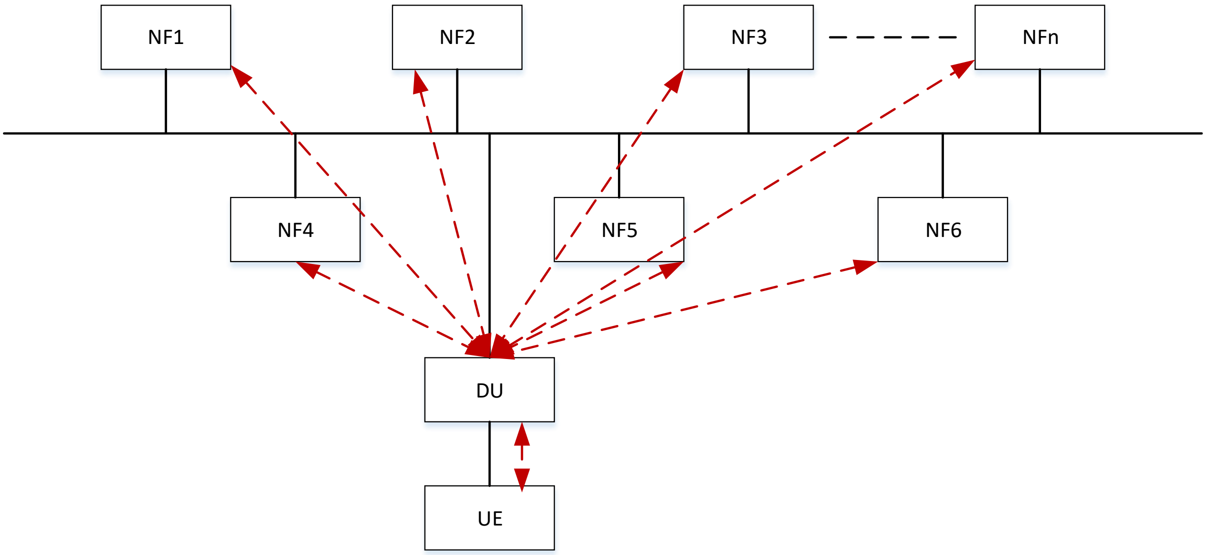


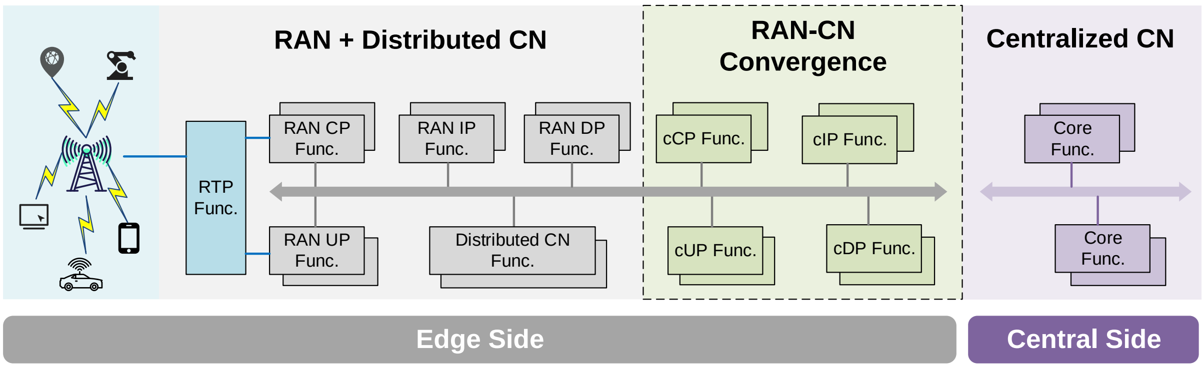
Figure 3-2: E2E SBA

The E2E SBA shows the following notable advantages:

- Efficient interaction with different NFs: RAN and CN NFs are not distinguished in the E2E SBA. All the NFs in the network communicate with each other via standardized service-based network APIs.
- Service-oriented communications: Each NF exclusively provides a specific service to other NFs through its service-based interface. For example, NF1 provides mobility management related service, while NF2 provides session management related service.
- Unified access between the RAN/CN and the UE: The UE communicates with all NFs in the network using control messages, which are delivered directly to the destination NF through the DU. For example, the UE directly communicates with a (potentially better integrated) single NF providing mobility management related service, and another NF providing session management related services.
- For the E2E SBA, the following key technologies should be considered:
  - Virtualization: The virtualization technology represented by virtual machines/containers could be used to provide elastic resource expansion and network functionality.
  - Microservices and service engines: In order to cope with the complexity of a large number of service calls and dependencies, the service engine technology could be considered to achieve the whole life cycle management of network services.
  - Service mesh: Service mesh enables the service instance to focus on the business logic itself.

Based on the above E2E SBA as shown in Figure 3-2, Figure 3-3 illustrates a reference RAN-CN converged architecture for 6G. The wireless network is deployed in a "central + edge" form, where the central side includes centralized core network functions, and the edge side includes RAN functions and distributed core network functions. Specifically, distributed core network functions may be logically separated from RAN functions or integrated with corresponding RAN functions.

New business and scenario requirements not only demand the network to provide basic connectivity services but also require intelligence services, edge computing services, high-precision sensing services, etc. The 6G edge-side network may add intelligence plane and data plane apart from the control plane and user plane. In Figure 3-3, the edge side adopts a service-based architecture. Real-time processing functions in the radio domain, with high requirements for real-time performance, might not initially be service-based. The four logical planes include multiple service-based NFs, described specifically as follows.



- RTP: Real-Time Processing
- cCP: converged Control Plane
- cUP: converged User Plane
- cIP: converged Intelligence Plane
- cDP: converged Data Plane

Figure 3-3: Reference RAN-CN converged architecture for 6G

The control plane primarily provides functions such as connection control, user plane control, and radio service management control. Connection control includes UE access control, radio bearer management, mobility management, radio resource management scheduling, radio connection security control, etc. User plane control functions are mainly responsible for session management, including the configuration, modification, release, etc., of session resources. Radio service management control is used for managing NFs in a SBA, such as providing service registration and deregistration, service discovery, inter-service communication, authentication verification, load management, etc.

The user plane supports packet processing, such as data encapsulation, forwarding, etc., and provides guarantees for data reliability and data security. Additionally, leveraging the advantages of SBA, the user plane can flexibly customize data processing strategies to meet diverse performance requirements.

The introduction of the data plane facilitates the logical separation of network control and data management. Data plane functions are responsible for static and dynamic data storage and processing, and they are exposed to the control plane, user plane, and intelligence plane through standard interfaces.

The intelligence plane provides AI-related functions, such as data modelling, model training, inference decision-making, knowledge graphs, etc. It serves the intelligence of the 6G network itself and meets the intelligence requirements of users and business applications.

RAN-CN convergence can occur in these four logical planes, depending on the scenarios and business characteristics. For example, in the control plane, RAN-CN convergence may be achieved in aspects such as UE access control, mobility management, session management, radio connection security control, etc.

For state management functionality, based on the description in Section 3.4, in the current architecture shown in Figure 3-4, the RAN and the AMF separately handle RRC state management and CM state management, respectively. The state management functionality in the CN can be offloaded to the RAN and converged with the corresponding RAN functionalities. In the converged network shown in Figure 3-5, the N2 interface connection is no longer required, and the distinction between RRC state and CM state is eliminated, enabling unified connection state management.

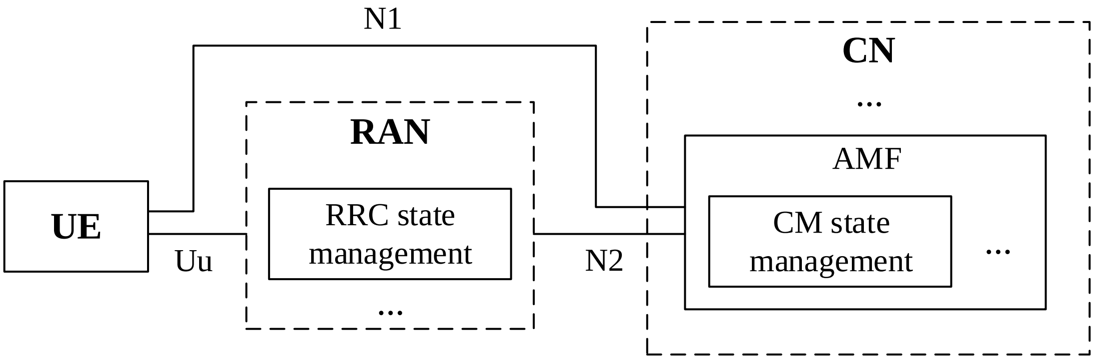


Figure 3-4 Current UE connection status management architecture

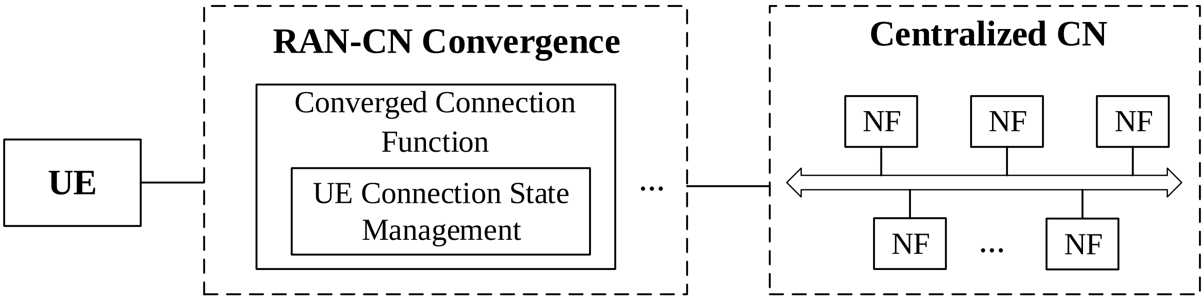


Figure 3-5 The UE connection state management architecture with RAN-CN convergence

In the user plane, RAN-CN convergence can ensure joint data processing. In the data plane, RAN-CN convergence can provide unified data storage and unified external exposure on the edge side. In the intelligence plane, RAN-CN convergence is beneficial for building AI cross-domain collaboration for data, models, and more. And there might also be some functions within the RAN and within in the CN that have not been converged, as shown in the grey areas of Figure 3-3, such as radio resource management in RAN and roaming in the CN.

3.5 Impact on the O-RAN architecture

The potential impacts on the current O-RAN architecture [4] due to RAN-CN convergence are explored in this section, encompassing aspects like architecture, protocols and interfaces, use cases, and signaling procedures.

1) Architectural Aspect

Virtualization and Cloudification: The virtualization and cloudification concepts of the O-RAN architecture can be extended in the RAN-CN convergence context. On the one hand, RAN-CN convergence can draw inspiration from and adopt the concepts of the O-RAN architecture to share resource pools [4] **Error! Reference source not found.** (such as computational resources). This can lead to more efficient resource utilization and dynamic allocation. On the other hand, O-Cloud's unified platform allows integrated monitoring, analytics, and management tools to oversee RAN and CN convergence. This simplifies troubleshooting, enhances operational efficiency, and provides insights for optimizing converged network performance.

Intelligent Control: The O-RAN architecture employs externalized RAN intelligence through Near-RT RIC and Non-RT RIC, realizing real-time, near real-time, and non-real-time control loops, as shown in Figure 3-6. RAN-CN convergence might involve the reconfiguration of intelligent functional architecture in RAN and CN domain. This might involve augmenting existing O-RAN functions (such as enhancing SMO for managing converged RAN-CN functionalities) and could potentially influence intelligent control across various levels of real-time responsiveness.

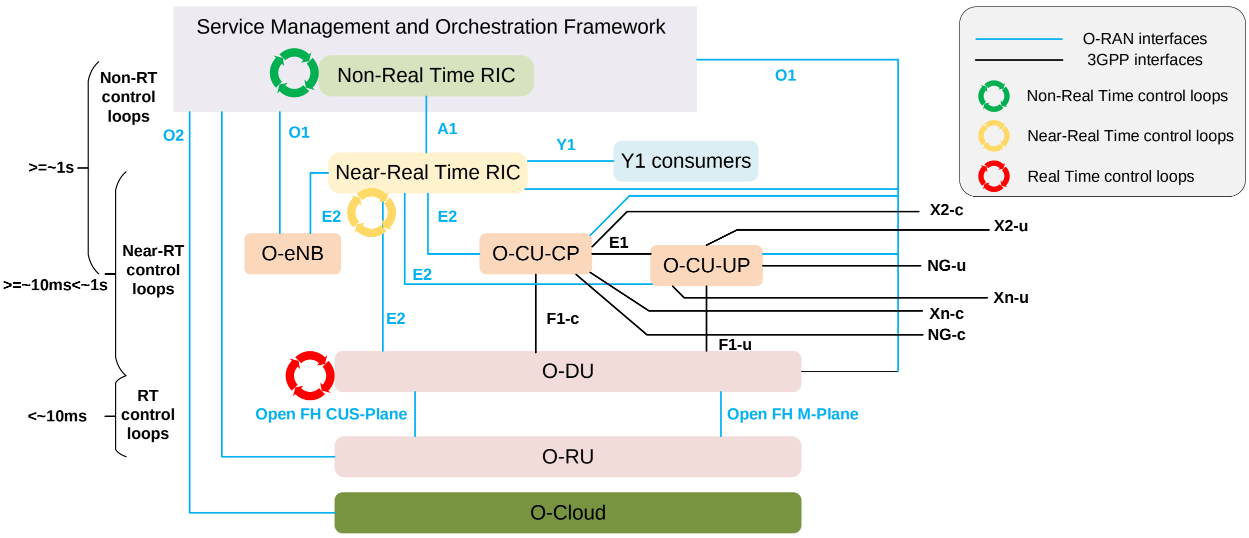


Figure 3-6: O-RAN Control Loops

2) Protocol and Interface Aspect

The RAN-CN converged functionalities deployed at the edge should engage with existing O-RAN functionalities like SMO, Non-RT RIC, and O-Cloud. Consequently, the corresponding O1, A1, and O2 interface protocols might necessitate expansion. Furthermore, if revolutionary innovations reshape the edge-side architecture, such as adopting a service-based structure to attain RAN-CN convergence, the creation of new protocols and interfaces would be imperative, including specialized protocols and standards-based protocols such as HTTP and QUIC.

3) **Use Case and Signalling Procedure Aspect**

Currently, O-RAN has conducted research and analysis on various use cases, such as context-based dynamic handover management for V2X, flight path-based dynamic UAV radio resource allocation, Quality of Experience (QoE) optimization, etc. The convergence of RAN-CN at the edge may further expand the scenarios and use cases supported by O-RAN, requiring a redesign of signalling procedures and call flows to incorporate and take advantage of the unified RAN-CN service-based models in existing and new use cases.

3.6 Impact on 3GPP architecture

1) **RAN Architecture and Radio Protocol Architecture**

5G RAN consists of gNB or ng-eNB, with gNB further divided into CU and DU nodes. Unlike the CN, the components of RAN are not divided according to functionalities. Moreover, RAN has a well-defined radio protocol architecture, including Physical (PHY), Medium Access Control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), Radio Resource Control (RRC), and Service Data Adaption Protocol (SDAP) sublayers, each with defined functionalities. RAN-CN convergence requires the reconstruction and/or integration of redundant functionalities of RAN and CN from a logical functional perspective, which may necessitate the redesign and adjustment of the RAN architecture and radio protocol architecture defined by 3GPP to meet these convergence requirements (e.g., restructuring at the highest level to incorporated a service-based RAN architecture). The CU functions may not exist in the form of a radio protocol stack but could be integrated with related CN functionalities. Whether to completely or partially restructure the architecture requires further discussion within the industry.

2) **Interface, Protocol and Procedure**

The convergence of RAN and CN may affect the functionalities of existing interfaces (N2 and N3) defined by 3GPP between RAN and CN. Additionally, it may introduce new interfaces and corresponding protocols, such as whether to introduce Service-Based Interface (SBI) on the RAN side to support the invocation of network functions or services after convergence. Furthermore, the convergence at the network function level may also impact the existing 3GPP signalling procedure logic, such as registration management, mobility management, session management, etc.

3) **Unified Resource Management and Scheduling Optimization**

RAN-CN convergence enables edge networks to comprehensively consider the optimization requirements of RAN and CN resource management and scheduling (e.g., computing resources, storage resources, bandwidth resources). Therefore, it requires the introduction of new resource management and scheduling strategies to achieve effective resource utilization and optimized allocation. However, existing 3GPP standards do not support the functionality of unified management and optimization. In some cases, it might be possible for O-RAN to extend the resource management mechanisms based on a simpler 3GPP RAN-CN architecture, but the basic architecture needs to be carefully defined in 3GPP (with minimal impact) to ensure nothing precludes these extensions.

4) **Security and Privacy Protection**

In existing 3GPP specifications, the CU terminates Access Stratum (AS) security, while the AMF terminates Non-Access Stratum (NAS) security. The AS security managed by CU ensures a secure connection between the UE and the RAN, mainly focus on encryption and integrity protection for AS signalling data and user data. Meanwhile, the NAS security handled by AMF is used for the ciphering and integrity protection of control plane signalling and user plane data between the UE and the AMF. RAN-CN convergence requires consideration of the division of security capabilities, which may affect existing security mechanisms. Additionally, for converged networks, there may be implications for the privacy protection of subscriber data.

4 Challenges of RAN-CN Convergence

For RAN-CN convergence, the following challenges may exist.

1) **Security and Privacy**

The converged network needs to consider the security and privacy protection of the entire edge-side network. Particularly, in the existing network, RAN cannot access CN-side user identity data and business data. Therefore, data security and privacy protection after RAN-CN convergence need careful consideration.

2) **Interoperability**

The 6G network may need to interoperate with 5G to achieve seamless handover and data interaction between heterogeneous networks. Since the RAN-CN convergence enhances the network architecture, radio protocols, and even terminal protocols, it poses a major challenge for interoperability and requires reasonable analysis and resolution.

3) **Standardization**

Firstly, RAN-CN convergence is more oriented towards edge-side industry scenarios, and it coexists with the traditional RAN-CN separation architecture for To Consumer (ToC) scenarios. Whether separate standards are needed for RAN-CN has not yet reached consensus in the industry. Secondly, for the 3GPP standardization organization, RAN and CN standardization are handled by independent working groups with distinct responsibilities. The division of responsibilities for standardizing RAN and CN independently (today) presents challenges for defining a unified RAN-CN converged architecture.

4) **Ecosystem**

In the current network deployment and device forms, RAN and CN are commonly independently designed and implemented. Achieving their convergence may require significant business motivation and collaborative efforts from various stakeholders in the industry chain.

5 Conclusion

This report summarizes preliminary research on RAN-CN convergence for the future development of 6G networks. The report introduces current RAN and CN architectures from O-RAN and 3GPP, then explores motivations and requirements for RAN-CN convergence, highlighting shortcomings of the current architecture and potential needs for future evolution. It defines RAN-CN convergence and analyses various use cases. By studying solution constraints and design principles, the report proposes a RAN-CN converged reference architecture based on an E2E SBA, considering a "central + edge" deployment model and adding an “intelligence plane” and an “data plane” evolved from the existing control plane and user plane. The impacts of convergence on existing O-RAN and 3GPP frameworks are analysed in terms of architecture, interfaces, protocol stacks, and signalling procedures. Challenges such as security, privacy, standardization, implementation, and ecosystem issues are also discussed.

In conclusion, this report serves as a foundational reference point for ongoing and future research endeavours focused on RAN-CN convergence within the context of 6G network architecture.

References

[1] 3GPP. TS 38.300, Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description.

[2] 3GPP. TS 38.401, Technical Specification Group Radio Access Network; NG-RAN; Architecture description.

[3] 3GPP. TS 23.501, Technical Specification Group Services and System Aspects; System architecture for the 5G System (5GS).

[4] O-RAN. O-RAN Architecture Description.

[5] Hexa-X. Deliverable D5.2 Analysis of 6G architectural enablers’ applicability and initial technological solutions.

[6] OPPO. White Paper, Zero-Power Communication.

[7] 3GPP. TS 28.104, Technical Specification Group Services and System Aspects; Management and orchestration; Management Data Analytics (MDA).

[8] 3GPP. TS 23.288, Technical Specification Group Services and System Aspects; Architecture enhancements for 5G System (5GS) to support network data analytics services.

[9] 3GPP. TS 32.130, Technical Specification Group Services and System Aspects; Telecommunication management; Network sharing; Concepts and requirements.

[10] 3GPP. TS 23.251, Technical Specification Group Services and System Aspects; Network Sharing; Architecture and functional description.

[11] 3GPP. TS 23.401, Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access.

[12] 3GPP. TS 23.502, Technical Specification Group Services and System Aspects; Procedures for the 5G System (5GS).

[13] O-RAN. White Paper. O-RAN Empowering Vertical Industry: Scenarios, Solutions and Best Practice.