

A Study on 6G Architecture for Unified UE Context Management

HyoChan Bang
Mobile Core Network Research
Section ETRI, Republic of Korea
bangs@etri.re.kr

MyungEun Kim
Mobile Core Network Research
Section ETRI, Republic of Korea
mekim@etri.re.kr

Namseok Ko
Mobile Core Network Research
Section ETRI, Republic of Korea
nsko@etri.re.kr

Abstract—In the 5G System (5GS), the Radio Access Network (RAN) and the Core Network (CN) are logically and functionally separated. Consequently, the User Equipment (UE) context generated during signaling procedures must be stored and managed independently by the UE, RAN, and CN. Furthermore, multiple network functions (NFs) within the CN are also required to maintain their own instances of the UE context to support signaling and session handling. This fragmentation of UE context across various functional blocks in 5GS necessitates additional control signaling and synchronization mechanisms, leading to increased signaling overhead and signaling latency. This paper analyzes the current UE context management and the associated synchronization challenges in the 5G architecture, and proposes a unified UE context management architecture. The proposed design enables shared access to UE context not only between the RAN and CN, but also among the various NFs within the CN. By facilitating efficient context sharing, the architecture reduces signaling overhead and signaling latency, thereby providing a scalable and low-latency foundation for emerging 6G applications.

Keywords— 6G, RAN-CN Convergence, UE Context, Service-Based Architecture (SBA), Service-Based Interface (SBI).

I. INTRODUCTION

The 5GS adopts a separated architecture with a functional separation between the RAN and the CN. The CN is designed based on the Service-Based Architecture (SBA), comprising multiple NFs. This innovative architecture enables key technologies such as network slicing, Multi-Access Edge Computing (MEC), and fine-grained Quality of Service (QoS) control. While functional decomposition provides flexibility and scalability, it also introduces significant complexity in call procedures and leads to the dispersion of User Equipment (UE) context information. As a result, redundant generation and processing of UE context are required between the RAN and the CN, as well as among individual NFs within the CN. Additionally, extra signaling overhead is incurred to maintain synchronization of the UE context across these entities. This distributed state management causes signaling latency due to duplicated procedures and fragmented context handling across network domains. The impact of this problem becomes more pronounced in 6G environments. Emerging 6G applications—such as Extended Reality (XR), Telerobotics, and Multimodal Streaming—are expected to require multiple PDU (Protocol Data Unit) Session with distinct QoS profiles simultaneously. Each PDU session may be mapped to a different network slice, resulting in concurrent UE registration, PDU Session establishment, and handover procedures for a single service. Consequently, UE context will be dispersed across an increasing number of functional blocks, inevitably requiring more complex synchronization and recovery mechanisms, leading to service delays. These challenges present significant constraints to achieving the ultra-low latency and high reliability targets of 6G networks. To address these limitations

inherent in the 5G architecture, this paper proposes a unified UE context management architecture. The proposed approach aims to reduce signaling overhead by providing an integrated and optimized method for managing UE context across the network. This proposed architecture is expected to improve scalability, reliability, and responsiveness, thus fulfilling the stringent requirements of next-generation mobile systems.

II. ARCHITECTURAL LIMITATIONS OF 5GS PROCEDURES

In the 5G mobile system, the RAN and the CN are logically and physically separated, operating as independent functional domains. As shown in Table I both the registration and PDU session establishment procedures involve duplicated and distributed management of UE and session state information across multiple network functions.

TABLE I. REVIEW OF THE DIVISIONAL FUNCTIONS OF 5GS

	RAN CU	AMF	SMF	PCF	AUSF	UDM
Registration	✓	✓			✓	
Session Est.	✓	✓	✓	✓		✓

A. Redundant UE Context Coordination in 5G

Call procedures in 5G networks involve coordinated interactions between the RAN CU (Central Unit) and various control plane entities in the core network, such as the Access and Mobility Management Function (AMF), Session Management Function (SMF), and Policy Control Function (PCF), following standardized procedures to establish network resources. During these procedures, each functional block is responsible for processing and sharing various types of UE context—including access state, authentication status, session state, and QoS parameters—as summarized in Table II. In particular, the RAN and the AMF must exchange and synchronize UE context over the N2 interface. This synchronization involves not only frequent signaling exchanges but also introduces potential bottlenecks, especially during error recovery or failure handling. The fragmented context and tight interdependence between entities increase signaling overhead and can significantly degrade overall system performance under complex or failure-prone scenarios.

B. Fragmented UE Context across NFs

The 5GS core network adopts a microservice-based architecture, where functions such as access, mobility, session, and policy control are implemented as independent network functions. While this design enhances the scalability and flexibility of each functions, it inevitably leads to the dispersion of UE context across multiple NFs, making context synchronization essential. For instance, during the PDU session establishment procedure, each NF—such as the AMF, SMF, and PCF—references and modifies UE context, including registration status, session state, and QoS policy. To maintain consistency, a significant amount of signaling is required to share these updates among the NFs. Such complex

This work was supported by the ICT R&D program of MSICT/IITP. [RS-2024-00405354, Development of Evolved SBA Framework and Core Technologies of Control/User Plane NFs]

propagation of UE context not only increases signaling overhead but also introduces synchronization delays or inconsistencies during fault recovery. As 6G networks demand higher agility and reliability, reducing the overhead caused by distributed UE context synchronization becomes a critical requirement.

C. Impact of RAN–Core Interface Heterogeneity

In 5G architecture, the RAN and CN communicate using NGAP(NG Application Protocol) over a synchronous messaging model, whereas CN NFs interact via the SBI, which employs HTTP/2 RESTful APIs and asynchronous messaging. This interface heterogeneity can introduce challenges in protocol conversion and coordination. The conversion between synchronous and asynchronous protocols increases protocol overhead and may disrupt processing consistency—particularly under high-concurrency scenarios such as parallel session establishments or inter-slice mobility. These issues pose critical limitations for ultra-low-latency 6G environments.

TABLE II. REDUNDANTLY MANAGED UE CONTEXT ACROSS NFs

State Information	Related NFs
UE Identifiers (SUPI, GUTI, IMSI)	AMF, SMF, UDM, PCF
UE Security Context	AMF, RAN-CU
RRC Status (RRC Connected/Idle)	RAN-CU, AMF
PDU Session ID & Status	SMF, PCF, RAN-CU
QoS Info (SDF, AMBR, etc.)	SMF, PCF, RAN-CU
N3 Tunnel Info (TEID, UP Path)	SMF, RAN-CU

III. PROPOSED ARCHITECTURE

The proposed architecture addresses key challenges in simplifying and enhancing the efficiency of signaling procedures required for mobile service establishment by introducing a converged RAN–CN framework. To solve these challenges, we explore how 5GS functions can be effectively integrated and restructured. Although the 5G standard architecture is designed to support a complete separation between the RAN and CN, in practice, numerous procedures involve frequent interactions between the two domains[1]. Consequently, the architecture inherently suffers from inefficiencies, including UE context desynchronization and redundant processing. As illustrated in Fig. 1, we propose shifting from a vertically disaggregated architecture to a procedure-oriented, horizontally integrated Architecture.

A. Converged RAN–CN Functional Blocks

Advancements in RAN disaggregation, virtualization, and cloudification have enabled the software-based implementation of the RAN-CU[2]. Building on this foundation, this paper proposed novel functional blocks that integrate RAN and CN control and user plane functions, as illustrated in Fig. 1. The control plane functions of the virtualized RAN-CU are integrated with the access management capabilities provided by the AMF through the Unified Access Control Function (UACF), while the mobility management functions are retained in the existing AMF. Similarly, the Unified Data Control Function (UDCF) integrates the user plane function (UPF) of the CN with the user plane components of the RAN-CU. Both UACF and UDCF maintain standard connectivity with the Distributed Unit (DU) via the conventional F1 interface, while connecting directly to other CN functions through the SBI. Since the virtualized RAN-CU can be deployed within the same cloud environment as the CN components, a flatter control plane

architecture becomes feasible. By eliminating the NGAP-based N2 interface between the RAN and AMF, UACF reduces redundant signaling operations previously performed in both domains, thereby minimizing control plane overhead and latency in procedures such as registration, session establishment, and mobility management.

B. Unified Context Management (UCM)

To address the fragmentation of UE context across multiple RAN and CN functions, this paper introduces a novel network function termed UCM (Unified Context Management). UCM is designed to centrally manage UE-related context information—such as security parameters, session attributes, QoS policies, and N3 tunnel data—that is dynamically generated throughout the call flow. Network functions store context in the UCM and may subscribe to updates when changes occur. Through the SBI extension up to the UACF in the RAN domain, context information can also be shared across access and core functions, enabling more efficient procedure execution. By merging UE context management, UCM reduces synchronization overhead among network functions and mitigates issues such as context inconsistency and recovery-related latency.

C. Harmonization of RAN and CN interfaces

The UACF exposes its functions via the SBI and directly interacts with other CN functions. By replacing the traditional point-to-point (P2P) RAN–CN interface with an SBI framework, the proposed architecture alleviates bottlenecks at interface transition points. Moreover, the asynchronous nature of SBI enables parallel execution of control-plane procedures, thereby reducing end-to-end service latency.

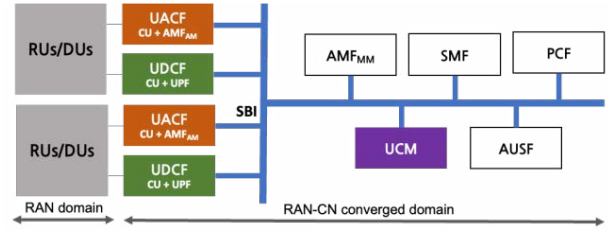


Fig. 1. Proposed Context Sharing RAN-CN Converged Architecture

IV. CONCLUSIONS

This paper analyzed the UE context synchronization challenges caused by the strict RAN–CN separation and the highly fragmented NFs of the CN. To address these challenges, this paper proposed an converged RAN–CN functional architecture in which RAN CU and Core Network NFs can share and utilize UE context in a unified manner. This approach suggests that a fundamental architectural shift may be required in future mobile systems, particularly in cloud-native 6G infrastructures. Future work will focus on implementing a prototype to validate the feasibility and performance of the proposed architecture.

REFERENCES

- [1] H. Watanabe, K. Akashi, K. Shima, Y. Sekiya, and K. Horiba, “A design of stateless 5G core network with procedural processing,” in Proc. 2023 IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom), Istanbul, Türkiye, Jun. 2023.
- [2] H. Harkous, A. Kak, A. Urie, H. Straulino, H. Wu, M. Laitila, H. T. Thieu, N. Choi, and T. Van de Velde, “Flat UP: Toward RAN-Core convergence for the 6G user plane,” IEEE Commun. Mag., vol. 63, no. 4, Apr. 2025.