Design of Resilient Mobile Core Network Architecture Inspired by Soft Guarantee Behavior

Masayuki Kurata, Ryu Watanabe, and Masaki Suzuki
KDDI Research, Inc., Saitama, Japan
{ma-kurata, ry-watanabe, and mak-suzuki}@kddi.com

***Abstract*—To initiate mobile communication, user devices must undergo authentication, authorization, and session setup, all handled by the mobile core network (MCN). Thus, an MCN outage can critically disrupt these procedures and degrade the user quality of experience (QoE). To address this, we propose the soft guarantee MCN (SG-MCN), an architecture inspired by the soft guarantee behavior from three-tier web systems. This technique achieves quick responses by bypassing bottleneck-prone data-tier processing and utilizing information from the application and presentation tiers. SG-MCN adopts this behavior to enhance resiliency while maintaining compliance with 3GPP security standards. Specifically, it verifies devices in a presentation-tier-equivalent function using a device-side-encrypted identifier and a precomputed identifier from the data tier. Mutual authentication is then performed using a precomputed authentication vector. Authorization and session setup are speculatively performed using enriched per-device contexts retrieved in previous procedures. Emulation results demonstrate that the SG-MCN sustains QoE scores close to those of the MCN under normal conditions, even in normal conditions with additional processing overhead and in outage conditions where data-tier functions are unavailable.**

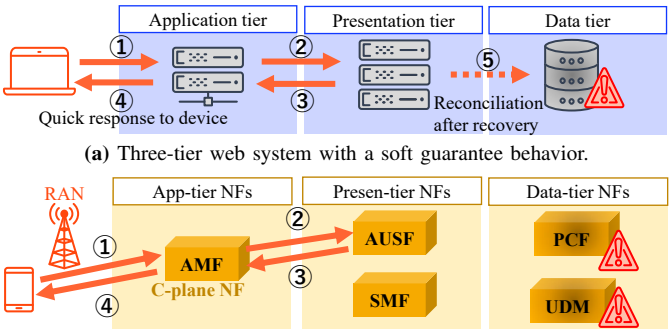
***Index Terms*—Mobile core network, Control plane, Resiliency**

I. Introduction

Ensuring mobile network resiliency is essential, as outages disrupt services and severely degrade user quality of experience (QoE). A key enabler of resiliency is the stability of control plane network functions (C-plane NFs) in the mobile core network (MCN) [1], which handles procedures essential for initiating mobile communication. In particular, the availability of C-plane NFs acting as databases (DBs) is critical, as they manage subscription and policy data to support user mobile plans. Failures or loss of reachability to these NFs can disrupt the smooth handling of the procedures. Such issues, caused by human operational errors or device specifications, have often been reported in commercial environments [2].

Although various approaches have been proposed to improve the resiliency of DB-related NFs, conventional solutions remain insufficient. A common method involves deploying backup instances at redundant sites [3]. However, network misconfigurations caused by human error have occasionally prevented these systems from functioning correctly, resulting in prolonged outages [4]. In addition, DB-related NFs are more difficult to distribute than other NFs due to strict synchronization requirements [5]. While AI-based techniques have been extensively studied to mitigate human error [6], completely eliminating such errors remains a significant challenge.

Motivated by these challenges, we propose a novel architecture, termed the soft guarantee MCN (**SG-MCN**). This is inspired by a speculative handling technique known as soft guarantee [7], originally developed for three-tier web systems.



(a) Three-tier web system with a soft guarantee behavior.

(b) Soft guarantee mobile core network (SG-MCN).

Fig. 1: Soft guarantee behavior in web systems and the MCN.

As illustrated in Fig. 1(a), when data-tier functions fail to operate correctly due to high load, a three-tier web system with a soft guarantee behavior bypasses data-tier processing. It then returns a response using information retained in the application and presentation tiers, which was previously retrieved from the data tier under normal operating conditions. This approach is referred to as speculative handling, as it generates responses without verifying the current state of the underlying databases. This realizes high QoE even in outage situations. The soft guarantee is considered applicable to flash deal scenarios that experience bursty access, such as Amazon Black Friday. For example, when a user attempts to purchase an item, its purchase availability is verified at the application and presentation tiers. After the data-tier functions recover, a reconciliation handling is performed, and if the result differs from the speculative response, an updated response is issued.

Although the soft guarantee behavior is promising for masking database-related NF failures in the MCN, its application presents significant challenges. This is because, unlike flash deals, where every user receives the same availability response regardless of individual attributes, the MCN requires user-specific processing. When authentication is speculatively performed, it involves reusing the authentication vector (AV) [8], i.e., a set of authentication elements, retained by the presen-tier NFs. This violates the security requirement of generating random challenge values in challenge-response authentication mechanisms [9]. Furthermore, in speculative authorization and session setup, the access permissions and session agreements that users are supposed to receive may not be properly applied. In other words, speculative procedure handling in

978-3-903176-75-1 ©2025 IFIP

the SG-MCN must ensure acceptable per-user accuracy while preserving security requirements.

Fig. 1 (b) illustrates the soft guarantee behavior implemented in the SG-MCN. To clarify its architectural analogy with web systems, MCN C-plane NFs [10] are classified into three logical tiers: app-tier, presen-tier, and data-tier NFs, corresponding to the application, presentation, and data tiers, respectively. Specifically, the app-tier NF is the access and mobility management function (AMF); the presen-tier NFs are the authentication server function (AUSF) and session management function (SMF); the data-tier NFs are the policy control function (PCF) and unified data management (UDM). Based on this tiered structure, the SG-MCN enables authentication, authorization, and session setup to be handled in a manner that masks outages of data-tier NFs, as outlined below.

Authentication: During failures of data-tier NFs, device authentication is handled by presen-tier NFs using an encrypted device identifier and an additional AV, both precomputed by the data-tier NFs under normal conditions. The encrypted identifier is generated using a pre-shared seed with the device and is expected to match the one generated by the device in its next authentication handling during outages. Consequently, even presen-tier NFs without decryption capabilities enable device verification by comparing the received identifier with the one stored from the data tier. The additional AV associated with its identifier is then used to execute a challenge-response authentication mechanism between the device and the SG-MCN, which prevents the reuse of the same AV.

Authorization and session setup: The app-tier NFs retain additional access- and session-related information in the enriched context to acceptably handle on a per-device basis. This extension aims to apply the device's policy and subscription data as long as possible, even during the outage. Once the affected NFs have recovered, reconciliation handling is performed to correct any inconsistencies, such as the application of a policy or subscription data that deviates from the intended mobile plan. This handling is essential for supporting flexible mobile plans, e.g., [11], that vary on an hourly or daily basis.

We implement a prototype of the SG-MCN using the open-source software free5GC [12] and UERANSIM [13], which emulate the MCN, and the device and radio access network (RAN), respectively. Emulation results show that the SG-MCN sustains QoE scores equal to or better than the MCN, even during outages, by bypassing the data-tier NFs. Moreover, under normal conditions with added processing overhead, the QoE scores also remain comparable to those of the MCN.

II. Background

A. Procedures for Initiating 5G Mobile Communication

To initiate 5G mobile communication, devices must undergo the Registration and Protocol Data Unit (PDU) Session Establishment procedures in the MCN [14]. Upon receiving a Registration request from a device, the MCN triggers a challenge–response-based authentication mechanism (see Sec. II-B). The AMF then completes authorization by repeatedly querying the PCF and UDM for access-related information. Afterward, PDU Session Establishment is performed. The SMF queries the PCF and UDM to retrieve session-related information and configures the UPF accordingly.

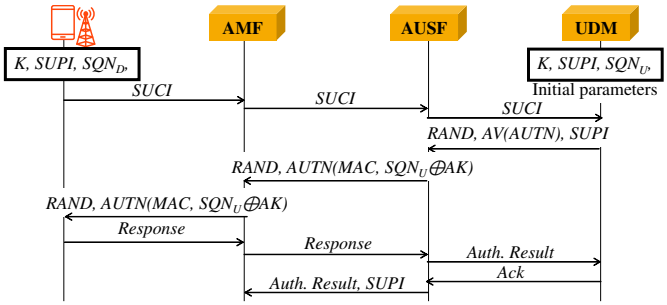


Fig. 2: Authentication signaling flows showing only elements relevant to this study.

Through Registration and PDU Session Establishment procedures, the AMF, AUSF, and SMF create access, security, and session contexts based on data retrieved from the PCF and UDM. When the device becomes unauthorized (e.g., when the device is turned off), the context stored in the SMF is deleted, while the context in the AMF is partially removed to release associated computing resources.

B. Overview of Standard Authentication Handling

Fig. 2 illustrates the 5G Authentication and Key Agreement (5G-AKA) mechanism and its key authentication elements, as stated in [8]. 5G-AKA is the primary method used for access authentication in 3GPP-compliant networks [15]. It is designed to (i) protect the subscription permanent identifier (*SUPI*) from exposure over the network, (ii) ensure forward secrecy by avoiding key reuse, and (iii) enforce a hierarchical key structure to isolate sensitive credentials.

The authentication begins with the user device transmitting a subscription concealed identifier (*SUCI*), which is derived by encrypting the *SUPI* using an ephemeral key. The ephemeral encryption key, ek_D, is derived through $\mathrm{KDF_{DH}}()$, a key derivation function based on Diffie–Hellman exchange, using the device-generated temporary private key sk_D and the UDM-provided public key pk_U:

$$ek_D = \mathrm{KDF_{DH}}(sk_D, pk_U) = \mathrm{KDF_{DH}}(sk_U, pk_D). \quad (1)$$

The device then encrypts the *SUPI* as follows:

$$SUCI = \{SUPI\}_{ek_D}, \quad (2)$$

where $\{\cdot\}_{ek}$ denotes an encryption operation with key *ek*. The resulting *SUCI*, along with pk_U, is sent to the UDM.

After that, the UDM decrypts the *SUCI* using the ek_D derived from its private key sk_U and the attached public key pk_D to recover the *SUPI*. It then retrieves the associated long-term key K and sequence number SQN_U. Subsequently, it generates an authentication vector (*AV*) using one-way functions based on K, SQN_U, and a random number *RAND*. During this process, an anonymity key (*AK*), message authentication code (*MAC*), and authentication token (*AUTN*) are also derived and embedded within the *AV*, which is then sent to the AUSF along with the *SUPI* and *RAND*.

The AUSF computes an expected response from the received *AV*, and forwards the *AUTN* and *RAND* to the device

via the AMF. The *AUTN* includes $SQN_U \oplus AK$ and *MAC*. The device computes the same *AK* from its own K and the received *RAND*, and recovers SQN_U by $(SQN_U \oplus AK) \oplus AK$. The device then compares the recovered SQN_U with its stored sequence number SQN_D; if the received value is not greater, the authentication is aborted to prevent potential replay attacks. Upon successful verification, SQN_D is updated to SQN_U.

The device next validates the *MAC* embedded in the *AUTN* using its own K and the received *RAND* to confirm the authenticity of the MCN. If successful, the device generates a response and sends it to the AUSF via the AMF. The AUSF compares this response with the expected value and, if they match, reports the authentication success to the UDM. These signaling processes complete the mutual authentication handling based on the challenge-response mechanism.

III. Soft Guarantee Mobile Core Network

As shown in Fig. 1(b), in the SG-MCN, the AMF cooperates with the AUSF and SMF to handle per-device requests based on soft guarantee principles during PCF and UDM failures. The AMF is mapped to the app-tier NF, as it is the first to receive signaling from devices via the RAN. The PCF and UDM, which manage subscription and policy data, constitute the data-tier NFs, while the SMF and AUSF, interfacing with both, are mapped to the presen-tier NFs.

The emergency procedure [14] enables authorization and session setup without PCF and UDM involvement but is limited to essential services (e.g., emergency calls) and relies on locally configured provisions in the AMF and SMF. In contrast, SG-MCN enables device authentication using pre-computed elements and supports speculative service delivery based on enriched pre-contexts during PCF and UDM failures.

A. Extended Authentication Handling

As illustrated in Figs. 3 and 4, unlike the MCN in Fig. 2, the device is provisioned with a shared long-term key K_S and a seed value *SEED*, while the UDM stores an additional seed value *SEED'* along with these parameters. The values of $SEED_D$ and $SEED_U$ are identical and updated for each authentication handling under normal conditions.

Normal-case Authentication: Fig. 3 illustrates authentication handling under normal conditions, where the challenge-response mechanism is executed following the generation of additional and extended authentication elements. In contrast to the standard authentication handling shown in Fig. 2, this process employs an extended authentication token (*eAUTN*), an additional authentication vector (*aAV*), another random number (*RAND'*), and an expected subscription concealed identifier (*XSUCI*), all generated by the UDM.

The use of *XSUCI* enables verification of *SUCI* without requiring decryption by the UDM. During UDM failures, no other NF provides decryption functionality. Delegating this functionality to other NFs would require transferring the UDM's private key, which violates the hierarchical key structure—a security principle that confines access to sensitive credentials to a single NF to ensure key isolation. To address

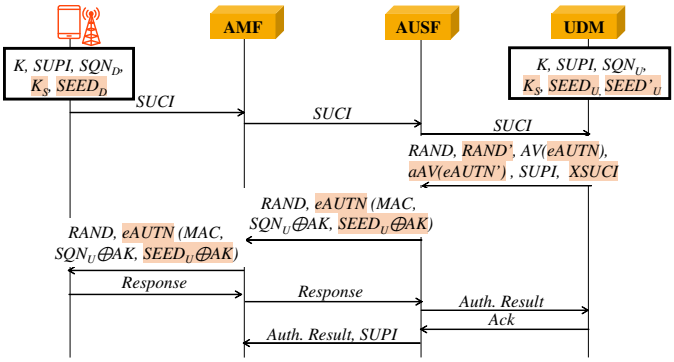


Fig. 3: Authentication signaling flows under normal SG-MCN conditions, with additional or extended elements highlighted.

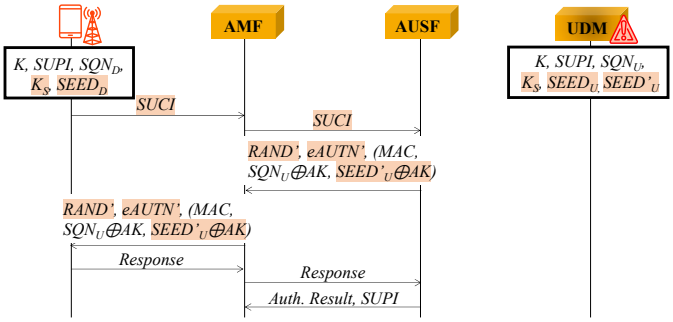


Fig. 4: Authentication signaling flows under outage SG-MCN conditions, with additional or extended elements highlighted.

this constraint, the UDM pre-encrypts the *SUPI* and generates the *XSUCI*, enabling verification of *SUCI* without runtime dependence on the UDM. Specifically, *XSUCI* is derived from a shared long-term key, K_S, and a synchronized seed value, $SEED_U$, both known to the device and UDM. The UDM first generates $SEED_U$ randomly and then computes the ephemeral encryption key ek using a hash-based key derivation function KDF() defined in the 3GPP specification [9] as follows:

$$ek = \text{KDF}(K_S, SEED_U) = \text{KDF}(K_S, SEED_D). \quad (3)$$

This key is then used to generate *XSUCI* following Eq. (2). Since K_S and *SEED* are protected in the same way as the private key of public key pairs denoted in Sec. II-B, the encryption key ek derived from them by using the KDF is also secure.

To ensure synchronization and prevent reuse, $SEED_U$ and $SEED_D$ are periodically updated with the device during each authentication handling. Accordingly, the *AUTN* is extended to *eAUTN*, which embeds $SEED_U$. If $SEED_U$ is included in *eAUTN* in plaintext, an interceptor who obtains *eAUTN* on the network and possesses K_S could reconstruct a valid *SUCI*, thereby compromising identity confidentiality. To prevent this, $SEED_U$ is encrypted in the same manner as SQN_U, using the anonymity key *AK*, and is included as $SEED_U \oplus AK$ in *eAUTN*.

In addition, the UDM generates an additional authentication vector (*aAV*) intended for use during UDM failures to ensure forward secrecy. To differentiate *aAV* from the previous authentication vector (*AV*), the UDM increments the sequence number as $SQN_U \leftarrow SQN_U + 1$ and generates a new random value *RAND'*. Without these updates, the resulting *aAV* would

be identical to the previous *AV*, thereby compromising forward secrecy. Moreover, on the device side, the received sequence number SQN_U is compared with the locally stored value SQN_D, and authentication is aborted if the received value is not greater. The extended authentication token in *aAV*, denoted as *eAUTN'*, includes the updated sequence number SQN_U and a newly generated seed value $SEED'_U$.

The generated authentication elements are transferred to the AUSF. The AUSF retains both *aAV* and *XSUCI* as related, and executes the challenge-response mechanism using the initially generated *AV*. Upon receiving *eAUTN* via the AMF, the device extracts $SEED_U$ by applying an exclusive OR operation with the *AK*, and sets $SEED_D$ to the extracted $SEED_U$. The remaining steps follow the standard authentication handling in Fig. 2, except for the report to the UDM.

Outage-case Authentication: As shown in Fig. 4, when the AMF detects a UDM failure, the authentication is handled in soft guarantee behavior. In this mode, the device applies the *SUPI* encryption method defined in Eq. (3), which uses the shared long-term key K_S and the synchronized seed value $SEED_D$, instead of the standard method in Eq. (1). The resulting encrypted identifier, *SUCI*, is sent to the AUSF.

Upon receiving the *SUCI*, the AUSF checks whether it retains the corresponding expected value, *XSUCI*. If a match is found, the AUSF initiates mutual authentication with the device using the associated authentication vector, *aAV*. The elements *RAND'* and *eAUTN'* from *aAV* are delivered to the device via the AMF, in accordance with standard procedures. Upon reception, the device performs the challenge-response mechanism shown in Fig. 2, except for the step involving result reporting to the UDM.

It is important to note that $SEED'_U$, extracted from *eAUTN*, is not utilized for authentication handling under the quasi-normal scenario in which the RAN, AMF, and AUSF are functioning properly. It is included solely to populate the corresponding field in *eAUTN*. If $SEED'_U$ were to be used, it would be in the rare case where the Registration or PDU Session Establishment procedures are re-executed during UDM failures. In such situations, the additional *XSUCI* and *aAV* retained by the AUSF are required to avoid reusing the same authentication challenge. However, as such re-executions are infrequent in successful cases [16], this work assumes the use of a single *XSUCI* and a single *aAV* per device to reduce the additional computational overhead on the UDM.

Using multiple *AVs* reduces DB-related NF load by avoiding repeated authentications [17]. In contrast, our method assumes complete DB-related NF unavailability and uniquely preserves *SUPI* encryption in such cases. The use of *XSUCI* is novel, as earlier generations (e.g., 3G, 4G) lacked *SUPI* encryption and exposed permanent identifiers in plaintext.

B. Extended Authorization and Session Setup Handling

Normal-case Authorization and Session Setup: The AMF performs standard authorization and session setup handling to retrieve subscription and policy data from the PCF and UDM. Unlike the behavior described in Sec. II-A, the access

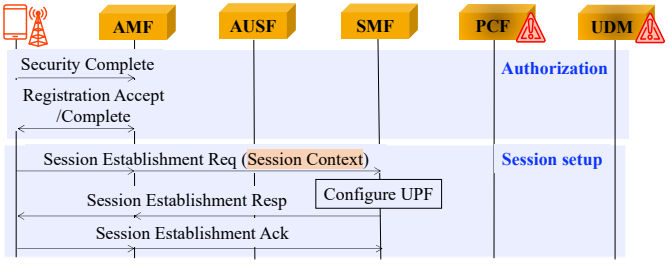


Fig. 5: Speculative authorization and session setup handling in the SG-MCN during the failures of the PCF and UDM.

context retained by the AMF is fully preserved even after the Deregistration procedure. Additionally, the SMF transfers the complete session context to the AMF in the "Session Establishment Resp" signaling, preventing its loss during the PDU Session Release procedure. As a result, the AMF retains both access- and session-related information.

Outage-case Authorization and Session Setup: In the soft guarantee mode, as shown in Fig. 5, the AMF bypasses queries to the PCF and UDM and speculatively completes the authorization process using its locally maintained access context. For session setup, the AMF provides the stored session context to the SMF, which then speculatively establishes the session based on this information.

Speculative handling poses no issue for static mobile plans (e.g., unlimited or fixed-volume), where service content remains stable, but may fail for dynamic plans such as daily, hourly, or usage-based billing. For instance, KDDI's "Povo" [11] offers high-speed, high-capacity access only during commuting hours. Applying outdated or unintended policies can degrade QoE, causing user dissatisfaction due to incorrect charges or service restrictions.

Upon recovery of the affected NFs, the AMF initiates reconciliation handling. It verifies whether the current subscription and policy configurations meet service requirements by querying the recovered NFs. If not, the configurations in the mobile network are updated accordingly. This ensures that the latest subscription and policy information is eventually aligned and correctly applied.

IV. Emulation-Based Evaluation

This section presents an emulation-based evaluation of the SG-MCN, focusing on authentication, authorization, and session setup handling. The SG-MCN is implemented using open-source software: free5GC [12], which emulates the MCN, and UERANSIM [13], which emulates the user device and RAN.

Emulation Environment: As shown in Fig. 6, the emulation environment is constructed using multiple virtual machines (VMs) created with VirtualBox, deployed across several physical servers. Each VM assigned to a C-plane NF is configured with 1 vCPU, pinned to a dedicated physical CPU core on the host server, and 4 GB of RAM. To emulate the user device and RAN, a VM is provisioned with 8 vCPUs and 16 GB of RAM. The emulation includes 5 RANs and 100 devices. Multiple RANs are used to mitigate a UERANSIM [13] limitation that causes signaling loss when

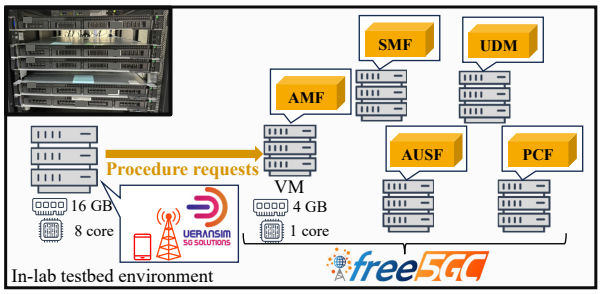
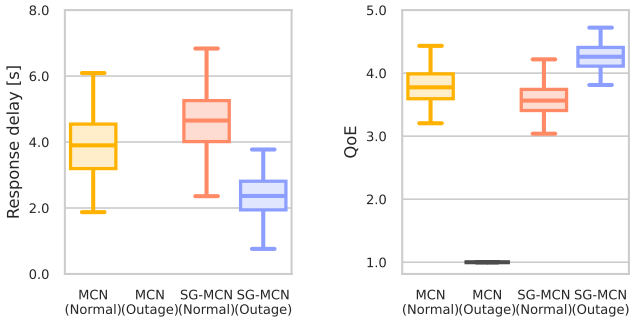


Fig. 6: Emulation environment.



(a) Response delays. **(b)** QoE scores.

Fig. 7: Emulation results of MCN and SG-MCN under normal and outage conditions.

100 devices simultaneously initiate Registration and PDU Session Establishment via a single RAN.

Experimental Setup: Before running MCN NFs, CPU and RAM caches are cleared to remove residual data and ensure consistent measurements. The MCN and RAN are then initialized, and 100 devices simultaneously initiate Registration and PDU Session Establishment. A 99% packet loss is imposed to emulate PCF and UDM failures, In the SG-MCN case, procedures are first run under normal conditions, followed by a device state reset and re-execution. Afterward, the same 99% loss is applied to evaluate SG-MCN under outage conditions.

Analysis on Response Delay: Figure 7a shows box plots of response delays, measured as the time from the start of Registration to the completion of PDU Session Establishment, based on device logs. For MCN (Normal), delays range from approximately 2 to 6 [s]. Under PCF and UDM failures, these procedures cannot be completed due to unreturned inquiries, so no box plot is shown for MCN (Outage). In contrast, SG-MCN (Outage) successfully completes both procedures by bypassing the failed NFs, resulting in significantly reduced delays compared to MCN (Normal). SG-MCN (Normal) exhibits higher delays than MCN (Normal), mainly due to the overhead of generating authentication elements such as *XSUCI* and *aAV*.

QoE-based Analysis: To evaluate whether the response delay affects user experience, we assess its impact on QoE using the model from [18], which estimates QoE based on delays in accessing web pages. Unlike video streaming, web access and procedures like Registration and PDU Session Establishment are discrete, single-request events. Thus, QoE is estimated as a function of response delay t:

$$\text{QoE} = 4.179 \cdot e^{-t/9.524} + 1.000. \quad (4)$$

Scores range from 1 (“bad”) to 5 (“excellent”) [18]. Figure 7b shows the QoE distribution across scenarios. In SG-MCN (Normal), scores mostly fall between 3 and 4. In MCN (Outage), all scores are 1, indicating severe degradation due to service unavailability.

These findings demonstrate that SG-MCN not only maintains QoE scores in the C-plane under normal conditions but also significantly mitigates service degradation during outages.

V. Conclusion and Future Work

This paper proposed the soft guarantee mobile core network (SG-MCN), an architecture inspired by soft guarantee behavior in three-tier web systems. To enable secure authentication under outage conditions, the 5G-AKA protocol is extended with mechanisms for *XSUCI* and *aAV*. The SG-MCN also leverages subscription and policy data in enriched contexts for per-device authorization and session setup, and applies reconciliation handling after recovery to restore consistency. The feasibility of the proposed approach has been confirmed through emulation using free5GC and UERANSIM.

Future work includes quantitative security analysis of extended authentication handling and corresponding enhancements. We also plan real-world experiments with actual equipment to validate the architecture in operational environments.

References

[1] R. Patil, Z. Tian, M. Gurusamy, and J. McCloud, “5g core network control plane: Network security challenges and solution requirements,” *Computer Communications*, p. 107982, 2024.
[2] D. Feng, S. Li, Z. Yang, Y. Xiang, J. Zheng, and X. He, “Research of deep learning and adaptive threshold based signaling storm prediction and top cause tracking,” *IEEE Access*, 2023.
[3] S. G. Kulkarni, G. Liu, K. Ramakrishnan, M. Arumaithurai, T. Wood, and X. Fu, “Reinforce: Achieving efficient failure resiliency for network function virtualization based services,” in *CoNEXT*, 2018, pp. 41–53.
[4] KDDI Corporation, “Outline About Communication Failure on July 2, 2022,” https://www.kddi.com/english/corporate/ir/ir-library/presentation/2023/220729-shougai/, July 2022.
[5] M. Kurata, A. Ikami, S. Itahara, and M. Suzuki, “Signaling storm mitigation by geographically distributed c-plane nf placement and routing,” in *IEEE NetSoft*, 2024, pp. 100–108.
[6] C. Benzaid and T. Taleb, “Ai-driven zero touch network and service management in 5g and beyond: Challenges and research directions,” *Ieee Network*, vol. 34, no. 2, pp. 186–194, 2020.
[7] Y. Niu, F. Liu, X. Fei, and B. Li, “Handling flash deals with soft guarantee in hybrid cloud,” in *IEEE INFOCOM*, 2017, pp. 1–9.
[8] 3GPP, “Security architecture and procedures for 5G System,” 3GPP, TS 33.501, 03 2025, version 19.2.0.
[9] ——, “Generic Authentication Architecture (GAA); Generic Bootstrapping Architecture (GBA),” 3GPP, TS 33.220, 03 2025, version 18.3.0.
[10] ——, “System architecture for the 5G System (5GS),” 3GPP, TS 23.501, 03 2025, version 19.3.0.
[11] KDDI Corporation, “povo2.0,” https://povo.jp/japan-sim/, 2025.
[12] “free5GC,” https://free5gc.org/.
[13] “Ueransim,” https://github.com/aligungr/UERANSIM.
[14] 3GPP, “Procedures for the 5G System (5GS),” 3GPP, TS 23.502, 03 2025, version 19.3.0.
[15] A. Koutsos, “The 5g-aka authentication protocol privacy,” in *IEEE EuroS&P*, 2019, pp. 464–479.
[16] J. Meng, J. Huang, Y. C. Hu, Y. Koral, X. Lin, M. Shahbaz *et al.*, “Characterizing and modeling control-plane traffic for mobile core network,” *arXiv preprint arXiv:2212.13248*, 2022.
[17] C.-K. Han, H.-K. Choi, J. W. Baek, and H. W. Lee, “Evaluation of authentication signaling loads in 3gpp lte/sae networks,” in *LCN*. IEEE, 2009, pp. 37–44.
[18] T. Tominaga, K. Sato, N. Yoshimura, M. Masuda, H. Aoki, and T. Hayashi, “Web-browsing qoe estimation model,” *IEICE Transactions on Communications*, vol. 100, no. 10, pp. 1837–1845, 2017.