CommIT Journal 20(2), 263–282, 2026

Performance Evaluation of Control Plane Signaling in 5G: A Novel Adaptive Software-Edge Architecture

Vivi Monita[1*], Naufal Hanan Lutfianto[2], Budi Prasetya[3], and Andrea Stevens Karnyoto[4]
[1]Smart City Information Systems Study Program, School of Applied Science, Telkom University
Jawa Barat, Indonesia 40257
[2,3]Telecommunication Engineering Study Program, School of Electrical Engineering,
Telkom University
Jawa Barat, Indonesia 40257
[4]Computer Science Department, BINUS Graduate Program – Master of Computer Science Program,
Bina Nusantara University
Jakarta, Indonesia 11480
Email: [1]monitavivii@telkomuniversity.ac.id, [2]naufalhananl@telkomuniversity.ac.id,
[3]budiprasetya@telkomuniversity.ac.id, [4]andrea.karnyoto@binus.ac.id

***Abstract*—The 5G core network faces increasing challenges in handling dynamic traffic growth, stringent latency requirements, and control-plane scalability due to its largely centralized processing model. Software-Defined Networking (SDN) improves network programmability, while edge computing reduces latency by decentralizing control and processing closer to users. However, existing SDN-based approaches for 5G cores still face limitations in dynamic signaling management, mobility handling, and resilience under varying traffic loads. The research proposes an Adaptive SDN-Edge 5G Architecture (ASE-5G) that integrates SDN programmability with edge-assisted control-plane coordination while preserving compatibility with thethe 3rd Generation Partnership Project (3GPP) service-based architecture. ASE-5G introduces rule-based and threshold-driven mechanisms to dynamically manage control-plane signaling, flow redirection, and controller failover without relying on GPRS Tunneling Protocol-User Plane (GTP-U) tunneling. The proposed architecture is evaluated using Objective Modular Network Testbed in C++ (OMNeT++) simulations and validated through analytical queueing models, focusing on control-plane signaling performance metrics, including latency, throughput, controller utilization, and failover behavior under different traffic conditions. Simulation results show that ASE-5G achieves low and stable signaling latency under moderate load, bounded throughput saturation under realistic protocol overheads, controlled resource utilization, and resilient controller failover behavior. A qualitative comparison with representative prior work indicates that ASE-5G extends existing SDN-based and edge-assisted 5G core designs by providing quantitatively validated performance insights while maintaining alignment with standardized 3GPP control procedures. These findings suggest that ASE-5G is a suitable and practical architecture for edge-centric and private 5G deployments requiring adaptive and resilient control-plane operation.**

***Index Terms*—5G Core Network, Software-Defined Networking (SDN), Edge Computing, Control-Plane Signaling, Adaptive SDN-Edge (ASE-5G)**

I. Introduction

THE fifth-generation (5G) mobile network has been envisioned as a transformative enabler of ultra-fast connectivity, massive device integration, and low-latency communication. Unlike its predecessor fourth-generation (4G), which has provided peak data rates of up to 1 Gbps for stationary users, 5G is designed to meet the exponentially growing demand driven by multimedia applications, mission-critical services, and the proliferation of Internet of Things (IoT) devices [1–5]. According to the Ericsson Mobility Report (2024) [6], global mobile data traffic reached the order of 157 exabytes per month in Q3 2024, and it was projected to approach 200 exabytes per month in 2025 nearly doubling in just a few years. This growth reflects not only the rising number of smartphone users but also the increasing consumption of high-bandwidth applications such as video streaming [7], cloud computing, and Augmented/Virtual Reality (AR/VR) [8]. Such trends pose significant challenges for mobile

Received: Oct. 10, 2025; received in revised form: Jan. 19, 2026; accepted: Jan. 19, 2026; available online: July 30, 2026.
*Corresponding Author

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,"Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture", CommIT Journal 20(2), 263–282, 2026.

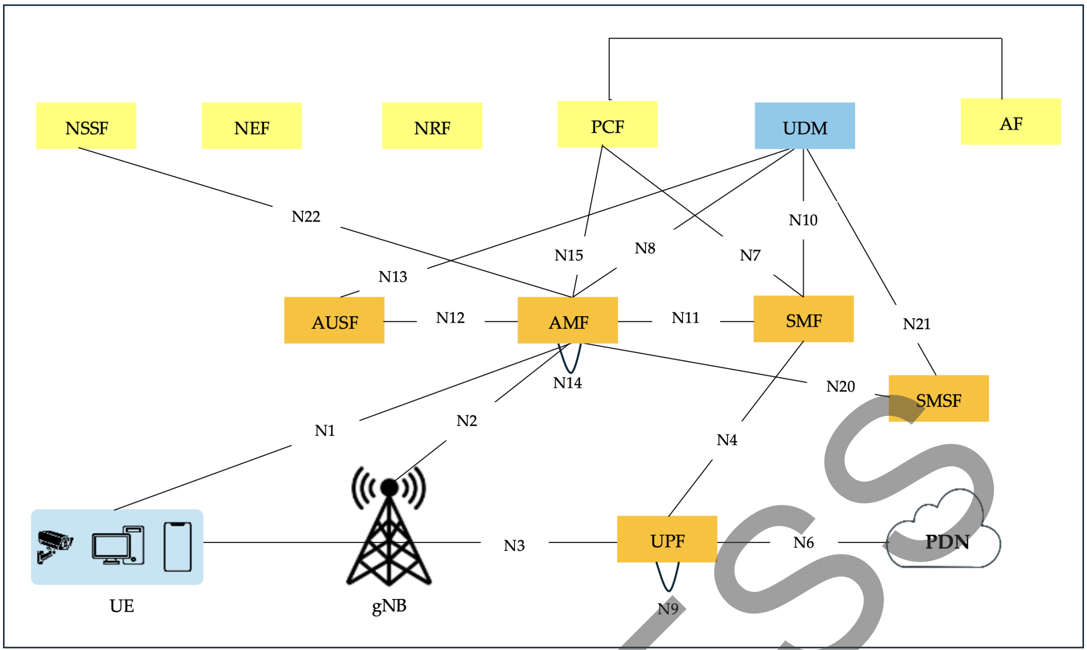

Fig. 1. 5G core network architecture based on the 3rd Generation Partnership Project (3GPP) service-based model. Note: Network Slice Selection Function (NSSF), Network Exposure Function (NEF), Network Repository Function (NRF), Policy Control Function (PCF), Unified Data Management (UDM), Application Function (AF), Authentication Server Function (AUSF), Access and Mobility Management Function (AMF), Session Management Function (SMF), Short Message Service Function (SMSF), User Equipment (UE), User Plane Function (UPF), and Packet Data Network (PDN).

operators in terms of scalability, latency reduction, and efficient resource management [4, 9, 10].

To address these challenges, the Next Generation Mobile Networks (NGMN) Alliance has outlined several 5G use cases, including enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communication (URLLC), and massive Machine-Type Communication (mMTC) [11–13]. Each use case imposes distinct requirements in which eMBB demands extremely high throughput, URLLC requires end-to-end latency as low as one millisecond, and mMTC emphasizes energy efficiency and large-scale connectivity [14]. Fulfilling these heterogeneous and increasingly stringent requirements simultaneously has proven difficult under conventional network architectures, particularly as traffic dynamics and service diversity continue to expand [15].

Beyond current 5G capabilities, research is increasingly shifting toward sixth-generation (6G) networks to support wider bandwidths, higher data rates, and intelligent services. 6G is expected to leverage upper mid-band and sub-terahertz spectrum to enable multi-gigahertz bandwidths and peak data rates reaching hundreds of gigabits per second (Gb/s) [16]. In addition, emerging concepts, such as Integrated Sensing and Communications (ISAC), introduce new control signaling and processing demands, particularly under high mobility and ultra-low latency conditions [17, 18]. These trends further emphasize the need for adaptive, low-latency, and resilient control mechanisms within the core and edge networks.

The 5G core network standardized by the 3rd Generation Partnership Project (3GPP) adopts a Service-Based Architecture (SBA), primarily defined in 3GPP Release 15 and enhanced in Release 16, which modularizes control and user plane functions [19–21]. Key components include the Access and Mobility Management Function (AMF), Session Management Function (SMF), Authentication Server Function (AUSF), Network Slice Selection Function (NSSF), Unified Data Management (UDM), and Policy Control Function (PCF), all of which interact through standardized interfaces to enable efficient network slicing and service differentiation [22–25]. The overall structure of the 5G system is illustrated in Fig. 1. It shows the interaction among the User Equipment (UE), Next-Generation Radio Access Network (NG-RAN), and core entities such as AMF, SMF, AUSF, and User Plane Function (UPF) through service-based interfaces. This modular design enhances scalability but also increases complexity in coordination, resource allocation, and signaling overhead as traffic and application diversity grow rapidly.

To manage this complexity, Software-Defined Net-

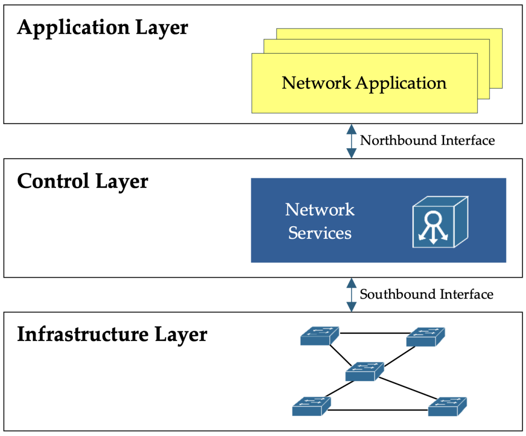

Fig. 2. Simple Software-Defined Networking (SDN) architecture with three layers: application, control, and infrastructure. The controller manages forwarding devices centrally.

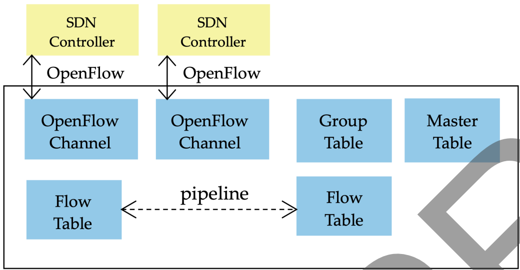

Fig. 3. Core elements of flow entry. Note: Software-Defined Networking (SDN).

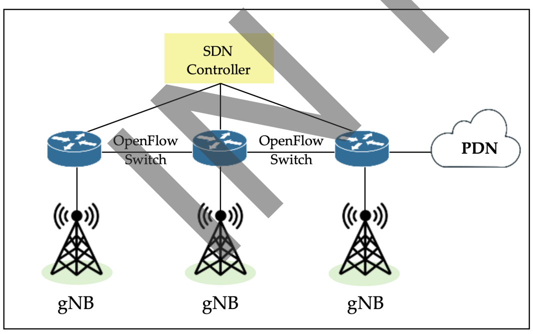

Fig. 4. Logically Centralized-Physically Distributed (LC-PD) control-plane design with logical centralization and physical distribution. Note: Software-Defined Networking (SDN) and Packet Data Network (PDN).

working (SDN) has emerged as a promising paradigm by decoupling the control-plane from the data-plane and enabling centralized programmability via open interfaces such as OpenFlow [26–28]. As illustrated in Fig. 2, the SDN architecture is typically composed of three layers. There are application, control, and infrastructure layers, which together enable programmable and centralized network management.

This separation improves flexibility, accelerates service deployment, and optimizes resource utilization, thereby reducing both capital and operational expenditures (CAPEX/OPEX) [26, 29–32]. At the infrastructure layer, OpenFlow switches manage data forwarding through flow tables, group tables, and meter tables, as illustrated in Fig. 3, enabling fine-grained traffic engineering and policy enforcement. Various SDN-based designs have been explored in mobile networks, such as SoftNet which combines centralized and distributed control for improved scalability [9], and the Logically Centralized-Physically Distributed (LC-PD) control-plane design, which coordinates multiple controllers to maintain a unified global network view for latency reduction and higher throughput. Figure 4 illustrates the LC-PD control-plane concept, where multiple distributed controllers collaborate under logical centralization, ensuring scalability and resilience against controller failures.

Complementing SDN, edge computing brings computation and control closer to end users, reducing latency and alleviating core network congestion [8, 12, 22, 32, 33]. Rather than relying solely on centralized data centers, edge nodes perform localized processing, thereby reducing latency, alleviating core congestion, and enhancing service reliability [32, 34, 35]. This architecture is particularly beneficial for time-sensitive applications, such as autonomous driving, remote surgery, and industrial automation, where even milliseconds (ms) of delay can jeopardize safety [1, 36, 37]. Previous research has demonstrated that combining SDN and edge computing improves traffic management and latency performance [1, 38–41]. For instance, relocating 5G core functions closer to the edge improves network responsiveness and efficiency [23, 42]. Nevertheless, many existing SDN-edge frameworks remain extensions of LTE-based designs and lack explicit alignment with the 5G SBA model [15, 43]. Moreover, legacy network management practices often rely on static configurations and manual or checklist-based evaluations, offering limited capability to adapt dynamically to fluctuating traffic loads, controller congestion, or signaling bursts [26, 44].

To address these limitations, this research proposes the Adaptive SDN-Edge 5G (ASE-5G) architecture. The adaptability in ASE-5G is achieved through rule-based and threshold-based control mechanisms, where key indicators, such as controller utilization, signaling queue occupancy, and registration or handover request rates, trigger dynamic control actions. These actions include adaptive flow redirection, controller failover, and signaling load redistribution across edge SDN controllers. Unlike prior SDN-based designs,

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

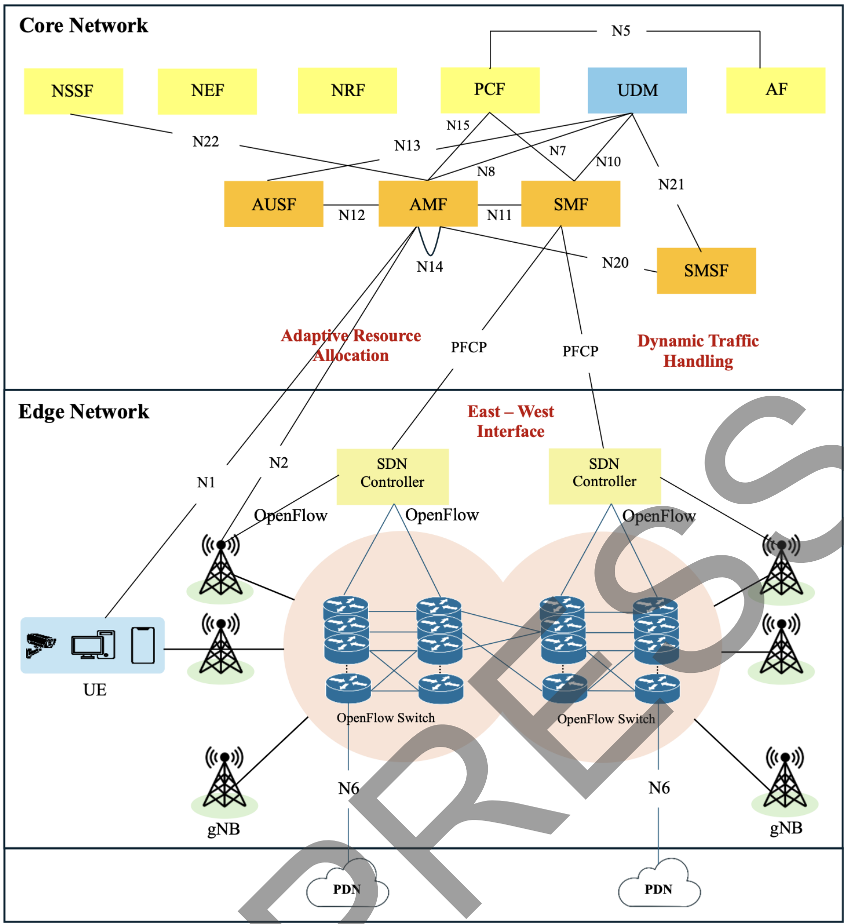

Fig. 5. ASE-5G architecture. Note: Software-Defined Networking (SDN), Network Slice Selection Function (NSSF), Network Exposure Function (NEF), Network Repository Function (NRF), Policy Control Function (PCF), Unified Data Management (UDM), Application Function (AF), Authentication Server Function (AUSF), Access and Mobility Management Function (AMF), Session Management Function (SMF), Short Message Service Function (SMSF), User Equipment (UE), User Plane Function (UPF), and Packet Data Network (PDN).

ASE-5G is explicitly aligned with the 3GPP service-based core model and introduces modified control-plane interactions to support seamless mobility and resilience. The performance of ASE-5G is evaluated using analytical queueing models and Objective Modular Network Testbed in C++ (OMNeT++) simulations, focusing on control-plane signaling procedures such as registration and handover under varying network loads. Comparative results demonstrate improvements in latency, throughput, resource utilization, and controller failover performance when compared with conventional SDN-based architectures. The contributions of this research are summarized as follows. First, the research proposes an ASE-5G architecture that integrates SDN and edge computing while preserving standard 3GPP release 15/16 control-plane functions. Second, the research introduces an adaptive control-plane mechanism based on rule-based and threshold-based triggers to dynamically manage signaling load and controller failover. Third, the research evaluates the proposed architecture through analytical modeling and OMNeT++ simulations to assess signaling latency, throughput, resource utilization, and resilience.

II. Research Method

A. Proposed Architecture

The proposed ASE-5G architecture integrates the programmability of SDN with the distributed intelligence of edge computing to address the limitations of conventional 5G core networks [13]. As illustrated in Fig. 5, the UPF is replaced by an SDN. While LC-PD and SoftNet mainly address controller scalability through distributed control, ASE-5G extends these concepts by being explicitly aligned with the 3GPP SBA control procedures and by introducing modified Packet Forwarding Control Protocol (PFCP)-based signaling between SMF and edge SDN controllers to support mobility-aware flow reconfiguration and failover without GPRS Tunneling Protocol-User Plane (GTP-U) tunneling [21].

Within the ASE-5G, the next-generation Node B (gNB) is redefined to function as an OpenFlow switch, supporting both its native control-plane roles and data-plane programmability [30]. Consequently, gNBs are required to support the OpenFlow protocol to interface with edge SDN controllers. It eliminates the need for tunneling protocols and avoids header encapsulation overhead, thereby improving data throughput. Unlike conventional deployments, gNBs in this architecture do not utilize the X_n application protocol to interconnect with one another. The X_n is the standardized interface between neighboring gNBs in 5G networks, primarily used to support handover signaling and coordination between adjacent gNBs. In ASE-5G, the SMF operates as an application-layer entity within the SDN network relaying forwarding information to the edge SDN controllers through the modified PFCP. It is important to clarify that ASE-5G is designed to preserve the standard 5G core control-plane procedures defined by 3GPP Release 15/16 service-based architecture, while selectively modifying the realization of the user plane [45]. The proposed architecture focuses on replacing the UPF implementation with an SDN-based data plane, without altering the logical roles of control-plane functions such as AMF and SMF.

The edge controllers in turn analyze and manage local resources, exchange state information across east–west interfaces, and install optimized forwarding rules into OpenFlow switches. Based on these rules, the switches forward packets according to flow table entries while continuously updating their state to the controllers. This adaptive interaction between SMF, edge SDN controllers, and OpenFlow switches establishes the foundation of ASE-5G, enabling improved latency reduction, dynamic resource allocation, and resilient traffic management under diverse network conditions.

B. Software-Defined Networking (SDN) Function Comparison

To further clarify the implementation of the proposed architecture, the ASE-5G framework introduces a new approach to user-plane processing by replacing the conventional UPF with an SDN-based data plane. In a conventional 5G network, the UPF acts as the central point for packet forwarding and mobility handling, especially for routing traffic when a device moves between different network segments [46]. In the ASE-5G design, this role is taken over by OpenFlow switches, which function as the main forwarding nodes in the data plane. These switches forward user traffic based on flow rules that are dynamically installed by the SDN controllers, allowing more flexible traffic management compared to the static behavior of a conventional UPF.

In the ASE-5G framework, the SDN-based data plane does not have a built-in mechanism to assign IP addresses, unlike a conventional UPF. For this reason, the architecture relies on the second and third methods IP assignment by the SMF or by external servers to ensure that every user device receives a valid address. This approach maintains compatibility with existing 5G core procedures while simplifying the data-plane design. Furthermore, even though the UPF has been replaced, the OpenFlow-based data-plane nodes still maintain full connectivity with external networks. It is achieved through forwarding rules configured by the edge SDN controllers, which determine how packets should be routed in and out of the network. These rules also allow the system to optimize traffic routing, improve responsiveness during mobility events, and maintain efficient connections between the 5G network and external data services [13].

Packet forwarding in ASE-5G is adaptive. The SMF sends Forwarding Action Rules (FARs) to the SDN-edge controllers through the modified PFCP. Upon receiving these rules, the SDN-edge controllers compute the optimal forwarding paths and configure the OpenFlow switches accordingly. Packets are then classified and processed according to performance-driven forwarding policies. The forwarding performance is evaluated using key metrics, including latency, throughput, resource utilization, and SDN controller failover. Based on these performance objectives, the SMF provisions the corresponding forwarding parameters into the flow rules, ensuring that packet forwarding is aligned with the intended network performance objectives.

Next, mobility management, particularly handover, is also redefined. In conventional designs, UPF generates end-marker packets to facilitate session release at the Source gNB (S-gNB) and packet reassembly at the Target gNB (T-gNB). In ASE-5G, this process is coordinated by the edge SDN controllers.

After computing the new forwarding path, the controller instructs the OpenFlow switch to release old flows and establish new ones. ASE-5G maintains functional equivalence to standard end-marker behavior through controller-assisted signaling and switch-level notifications. Instead of relying on dedicated end-marker packets in the data plane, the completion of flow termination is confirmed using OpenFlow mechanisms such as `OFP_FLOW_REMOVED` messages, which explicitly notify the controller when a flow expires or is removed. This approach preserves ses-

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

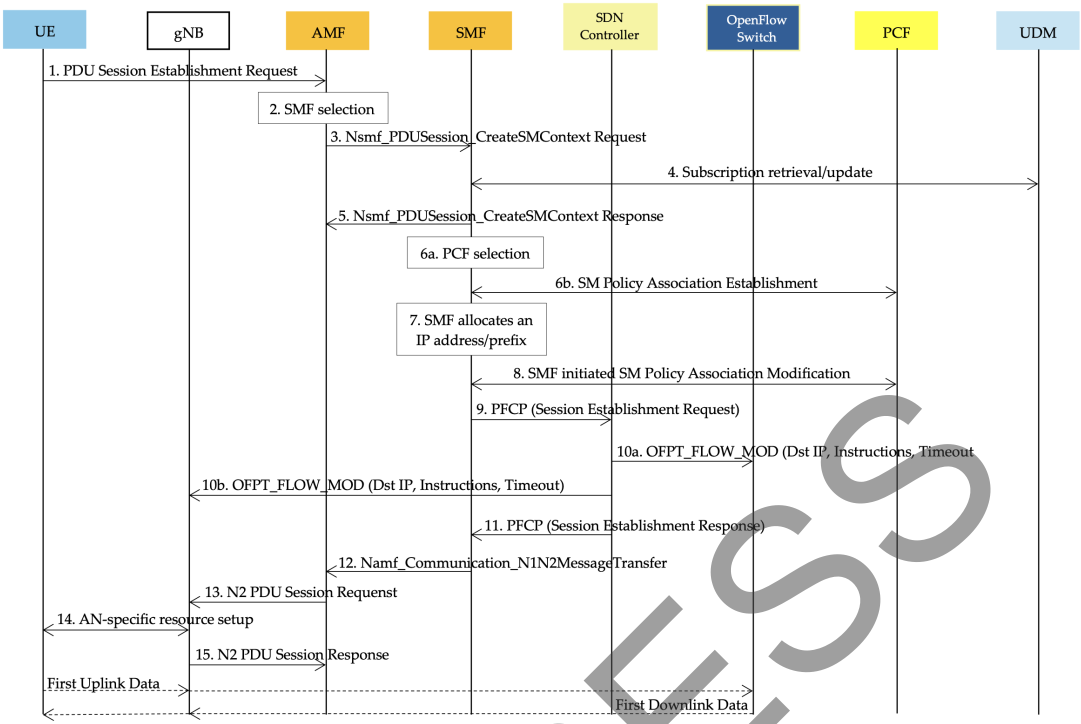

Fig. 6. Protocol Data Unit (PDU) session establishment procedure in ASE-5G architecture. Note: Adaptive SDN-Edge 5G (ASE-5G), User Equipment (UE), Next Generation Node B (gNB), Access and Mobility Management Function (AMF), Session Management Function (SMF), Edge Software-Defined Networking (SDN) Controller, OpenFlow (OF), Packet Forwarding Control Protocol (PFCP), Policy Control Function (PCF), Unified Data Management (UDM), Authentication Server Function (AUSF), Packet Data Unit (PDU).

sion continuity while reducing signaling overhead and simplifying data-plane operations. Flow entries are further reinforced with timeout mechanisms, allowing OpenFlow switches to autonomously expire stale flows or delete them upon explicit instructions (e.g., `OFPFC_DELETE` messages from controllers). These mechanisms directly contribute to performance metrics by reducing handover latency and improving resilience against failures.

C. Operational Procedures in ASE-5G

The Protocol Data Unit (PDU) session establishment procedure in ASE-5G is illustrated in Fig. 6. The AMF forwards the UE registration area information to the SMF through the `Nsmf_PDUSession_CreateSMContext` message. This information is then relayed to the Edge SDN controller, which calculates and determines the optimal forwarding path between the gNB and the OpenFlow switch. The SMF also communicates session establishment to the Edge SDN controller via a modified PFCP. The controller subsequently updates the flow entries on the OpenFlow switch to configure both the downlink (DL) and uplink (UL) paths. Unlike conventional networks, where DL and UL are established separately, ASE-5G configures both simultaneously. The controller then responds to the SMF through the modified PFCP to finalize the procedure.

The UE-triggered service request procedure is shown in Fig. 7. The SMF transmits a message to the Edge SDN controller to initiate a new session. The controller updates the flow entries on the OpenFlow switch and establishes both DL and UL forwarding paths. The controller subsequently sends a response to the SMF to confirm the session establishment. The network-triggered service request procedure is depicted in Fig. 8. When the OpenFlow switch receives a DL packet that it cannot process, it transmits a `PACKET_IN` message to the Edge SDN controller. The controller forwards the packet to the SMF, while the SMF temporarily buffers the packet until new forwarding rules are established.

Inter-UPF handover ASE-5G preserves the intent of the standard N2 handover with UPF reallocation while adapting it to an SDN data plane, as illustrated in Fig. 9. Instead of relocating a centralized UPF and maintaining GTP-U tunnels, ASE-5G updates forwarding at the relevant edge SDN controllers and OpenFlow switches using controller-assisted signaling

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

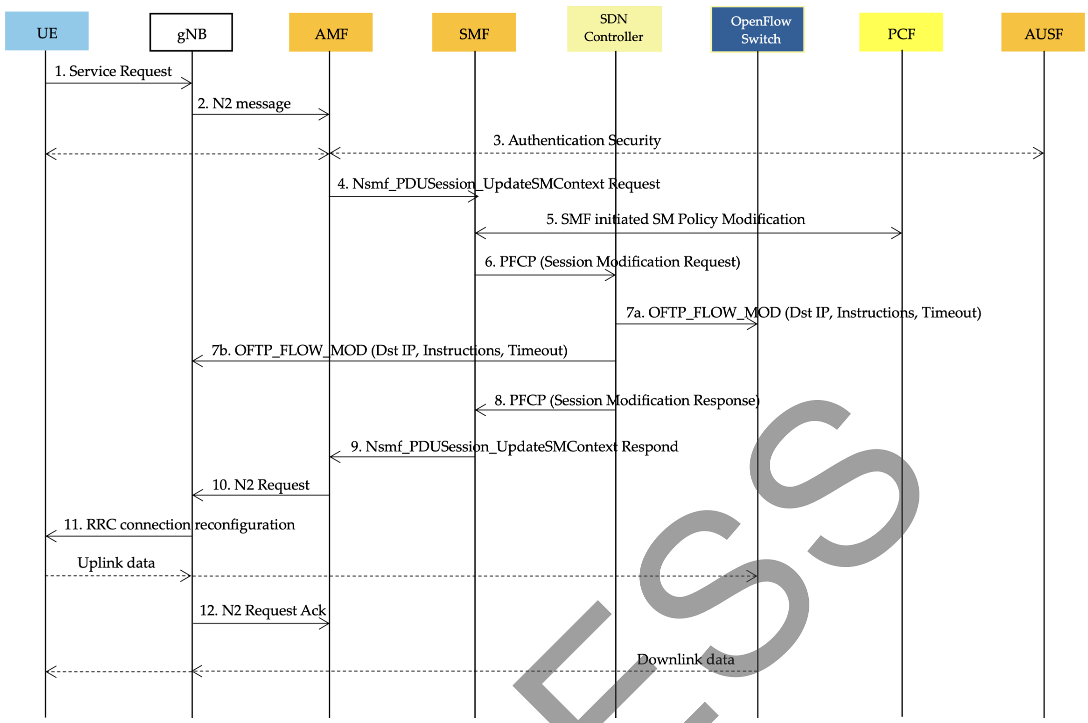

Fig. 7. Handover procedure in ASE-5G showing the role of edge Software-Defined Networking (SDN) controllers and OpenFlow switches. Note: Adaptive SDN-Edge 5G (ASE-5G), User Equipment (UE), Next Generation Node B (gNB), Source gNB (S-gNB), Target gNB (T-gNB), Access and Mobility Management Function (AMF), Session Management Function (SMF), Edge Software-Defined Networking (SDN) Controller, OpenFlow (OF), Packet Forwarding Control Protocol (PFCP), Packet Data Unit (PDU).

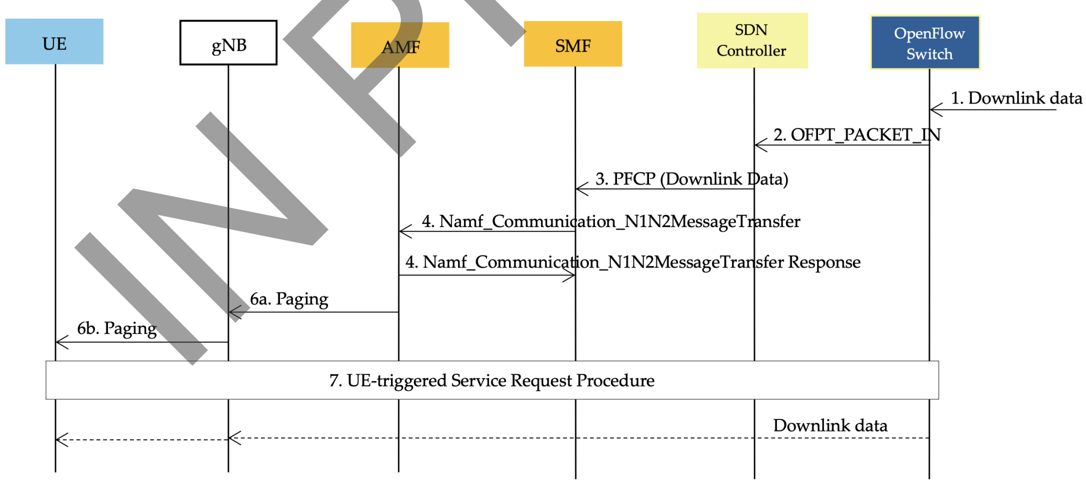

Fig. 8. ASE-5G procedure for handling network-triggered service requests. Note: Adaptive SDN-Edge 5G (ASE-5G), User Equipment (UE), Next Generation Node B (gNB), Session Management Function (SMF), Edge Software-Defined Networking (SDN) Controller, OpenFlow (OF), Packet Forwarding Control Protocol (PFCP), and Packet-In (`PACKET_IN`).

and switch notifications. The source-side controller detects the mobility event, coordinates with the target-side controller, and installs/updates flow rules to steer the UE traffic to the new path. The SMF still acts as the session anchor for policy and mobility decisions, but its actions are translated into SDN rule updates (via the modified PFCP messaging used in the prototype). It yields comparable handover behavior at the service level (session continuity, path update, and state synchronization) while reducing tunneling overhead and enabling edge-local control [20, 26, 30].

The intra-UPF handover procedure in ASE-5G is

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

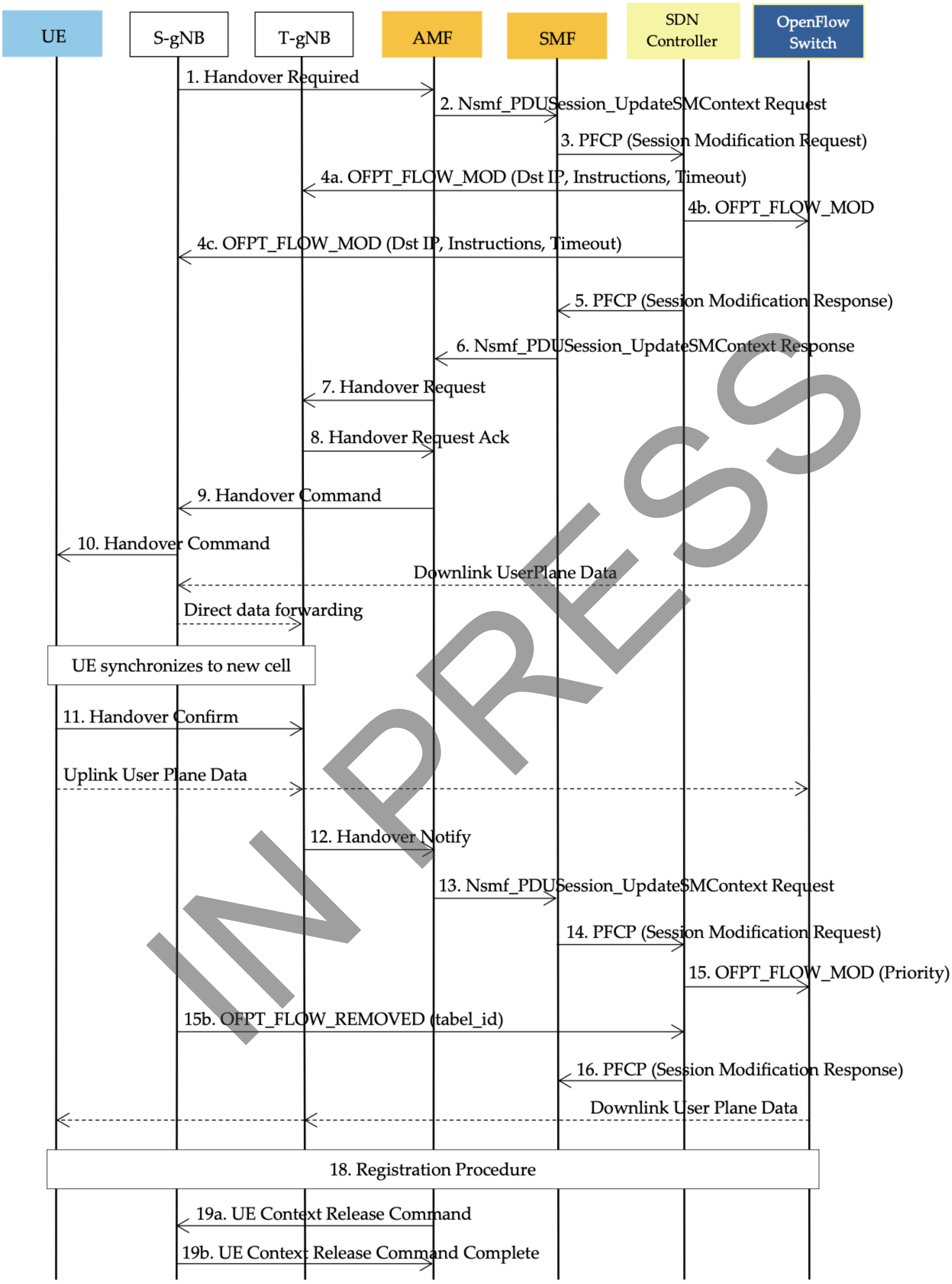

Fig. 9. ASE-5G workflow for inter-User Plane Function (UPF) handover. Note: Adaptive SDN-Edge 5G (ASE-5G), Source Next Generation Node B (S-gNB), Target Next Generation Node B (T-gNB), Access and Mobility Management Function (AMF), Session Management Function (SMF), Edge Software-Defined Networking (SDN) Controller, OpenFlow (OF), Packet Forwarding Control Protocol (PFCP), User Plane Function (UPF), OpenFlow Flow Modification (`OFPT_FLOW_MOD`), and OpenFlow Flow Removed (`OFPT_FLOW_REMOVED`).

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

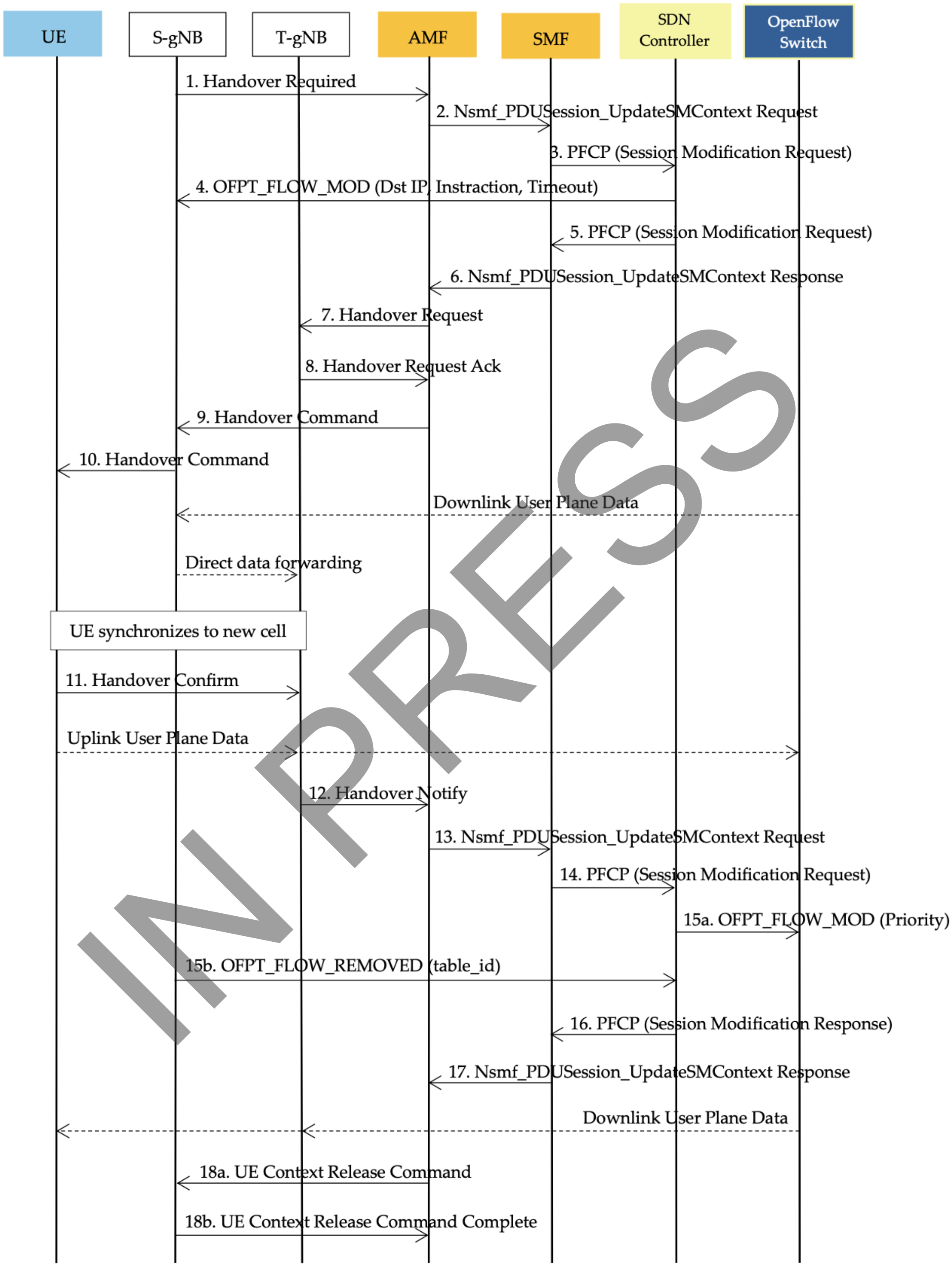

Fig. 10. ASE-5G workflow for intra-UPF handover. Note: Adaptive SDN-Edge 5G (ASE-5G), Source Next Generation Node B (S-gNB), Target Next Generation Node B (T-gNB), Access and Mobility Management Function (AMF), Session Management Function (SMF), Edge Software-Defined Networking (SDN) Controller, OpenFlow (OF), Packet Forwarding Control Protocol (PFCP), User Plane Function (UPF), OpenFlow Flow Modification (`OFPT_FLOW_MOD`), and OpenFlow Flow Removed (`OFPT_FLOW_REMOVED`).

shown in Fig. 10. During the initial PDU session creation, the forwarding path between the gNB and the OpenFlow switch is already set up to support normal data transmission. To maintain service continuity during handover and avoid packet loss, the Edge SDN controller dynamically configures an indirect data path from the S-gNB to the T-gNB. This additional path allows user traffic to continue flowing smoothly while the handover takes place, ensuring that no packets are dropped and the connection remains stable until the new direct path is fully established.

D. Analytical Model of the Proposed Architecture

The analytical modeling of the ASE-5G architecture is developed using the Jackson network framework, which is widely applied for analyzing multi-node communication systems. Under this framework, the system is assumed to operate in steady state, so the performance evaluation is independent of time. In addition, the arrival and service processes at each network node are assumed to be mutually independent and follow a Poisson distribution. These assumptions allow every processing node in the system including the gNB, SDN switch, and SDN controller to be represented as an M/M/1 queue. As a result, the arrival rate at each node can be computed analytically, and the overall delay performance can be derived using standard queueing theory.

Although the ASE-5G architecture adopts an LC-PD control-plane model with multiple distributed edge SDN controllers, the analytical model represents the control plane as a single logical controller. This abstraction models the collective behavior of multiple edge controllers as an aggregated service entity with equivalent processing capacity, which is sufficient to capture the average control-plane delay and throughput. Such an abstraction is commonly used in queueing-based performance analysis to simplify mathematical formulation while preserving essential system dynamics. As long as the distributed controllers operate under similar configurations and load-sharing mechanisms, the aggregated Jackson network provides a valid approximation of the ASE-5G control-plane behavior.

The analytical model focuses on the UE registration procedure, which is used as a representative proxy for evaluating handover delay and throughput in ASE-5G. Both registration and handover involve similar control-plane interactions, including signaling exchanges among the gNB, SMF, and SDN controllers, as well as forwarding rule installation and update processes. Consequently, the dominant delay components in both procedures are governed by comparable queueing and processing behaviors at the involved network entities. Therefore, the registration-based model provides meaningful insights into the relative delay and throughput performance of ASE-5G during mobility-related events, including handover scenarios.

Next, M/M/1 queueing model is adopted to enable tractable analytical derivation and to provide baseline insights into the control-plane behavior of ASE-5G. While real 5G traffic may exhibit bursty or self-similar characteristics, the focus of this analytical model is on control-plane signaling procedures, such as UE registration and handover, which are commonly approximated as Poisson processes in control-plane analysis. More complex queueing models such as G/G/1 can capture traffic burstiness more accurately. However, they significantly increase analytical complexity and reduce analytical transparency. Therefore, M/M/1 is employed as a first-order approximation to highlight the relationship between system parameters and performance metrics. The potential impact of bursty traffic is acknowledged as a limitation of the analytical model and is partially mitigated through the complementary discrete-event simulation presented in the following section.

The queueing structure of the ASE-5G system is illustrated in Fig. 11. Packets enter the system at the gNB with an initial external arrival rate denoted as λ_0. From the gNB, a fraction of packets is forwarded to the SDN switch with probability q_{bs}. At the switch, packets may either be processed directly or forwarded to the SDN controller with probability q_{sw}. The controller processes the incoming packets and returns a portion of them to the switch with probability q_c, forming a feedback loop commonly observed in SDN-driven architectures.

Based on this flow structure, the total arrival intensity at each node can be expressed analytically. Equation (1) defines the arrival rate at the gNB (λ_{bs}), which accounts for both the initial arrival rate λ_0 and the portion of traffic returning from the SDN switch. Then, Eq. (2) represents the arrival intensity at the SDN switch (λ_{sw}), which is influenced by the packets forwarded from the gNB and the packets looped back from the controller.

$$\lambda_{bs} = \lambda_0 + (1 - q_{sw})\lambda_{sw}, \quad (1)$$

$$\lambda_{sw} = q_{bs}\lambda_{bs} + q_c\lambda_{sw}. \quad (2)$$

Equation (3) describes the arrival rate at the SDN controller (λ_c), which depends on the fraction q_{sw} of messages forwarded from the switch. By substituting Eq. (3) into Eq. (2), the resulting closed-form expression for the switch arrival rate is obtained in Eq. (4). Next, substituting Eq. (4) into Eq. (1) yields

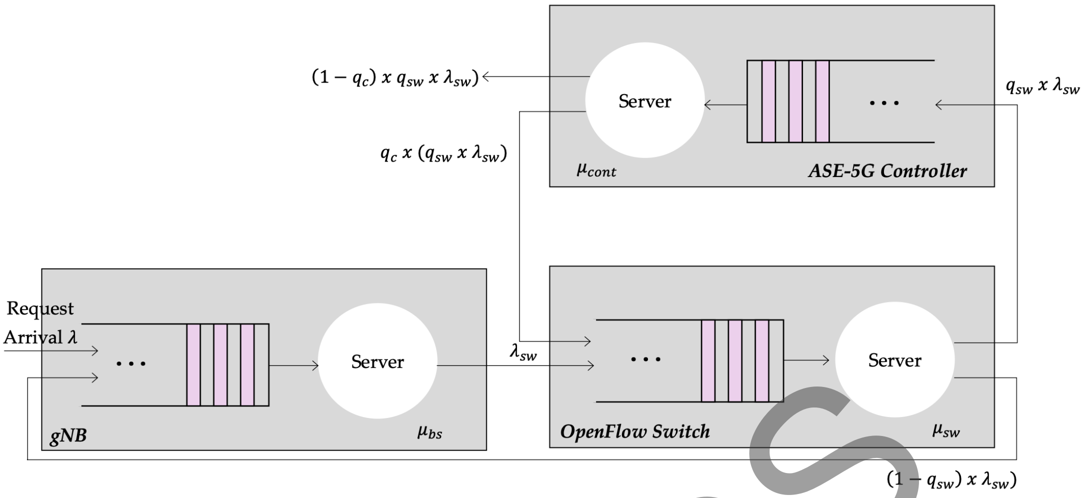

Fig. 11. Queueing model of the proposed ASE-5G architecture. The notation used is shown in Table I.

the expression for the gNB arrival rate, as shown in Eq. (5). Then, using Eqs. (4) and (5), the arrival rates for the switch and controller are derived in Eqs. (6) and (7).

$$\lambda_c = q_{sw}\lambda_{sw}, \tag{3}$$

$$\lambda_{sw} = \frac{q_{bs}}{1-q_c q_{sw}}.\lambda_{bs}, \tag{4}$$

$$\lambda_{bs} = \lambda_0 \frac{1-q_c q_{sw}}{1-q_c q_{sw}-q_{bs}(1-q_{sw})}, \tag{5}$$

$$\lambda_{sw} = \lambda_0 \frac{q_{bs}}{1-q_c q_{sw}-q_{bs}(1-q_{sw})}, \tag{6}$$

$$\lambda_c = \lambda_0 \frac{q_{bs}q_{sw}}{1-q_c q_{sw}-q_{bs}(1-q_{sw})}. \tag{7}$$

With these arrival rates, queueing-delay analysis can be performed. Since each node is modeled as an M/M/1 system with service rates μ_{bs}, μ_{sw}, and μ_c, the average waiting time at each node is given by $1/(\mu - \lambda)$. Thus, the total end-to-end delay for a registration request in ASE-5G is expressed in Eq. (8) as the sum of the mean delays at the gNB, switch, and controller. The formula is as follows:

$$E[D_{\text{TOT}}] = \frac{1}{\mu_{bs}-\lambda_{bs}} + \frac{1}{\mu_{sw}-\lambda_{sw}} + \frac{1}{\mu_c-\lambda_c}. \tag{8}$$

This analytical formulation demonstrates that the total end-to-end delay in the proposed ASE-5G architecture is jointly determined by the arrival rates and service capacities of the gNB, OpenFlow switch, and Edge SDN controller. As the traffic arrival rate at any network entity approaches its corresponding service rate, the queuing delay increases significantly due to congestion effects. Conversely, efficient traffic-forwarding probabilities and adequate processing capacities enable the system to maintain low latency, high service stability, and efficient control-plane signaling performance. Therefore, the proposed queueing model provides a quantitative basis for evaluating the scalability and latency performance of the ASE-5G architecture under varying traffic conditions. The notation and parameter definitions used in the analytical queueing model are summarized in Table I.

III. Results and Discussion

A. Results of Network Simulator

This research evaluates the performance of the ASE-5G architecture using key metrics, namely latency, throughput [27], resource utilization, and SDN controller failover. These metrics reflect essential aspects of network efficiency, stability, and resilience in 5G systems. All simulations are conducted with a fixed duration of 10 minutes, which is sufficient to capture the steady-state behavior of the system under different traffic loads. Each scenario is repeated 30 times, and the reported results represent the average values, reducing the impact of random variations and improving result reliability.

B. Comparative Performance Analysis with Standard 5G Core and SoftNet

In terms of throughput, ASE-5G reaches saturation at approximately 1,000 packets/s in OMNeT++, whereas the analytical queueing model predicts an idealized saturation point of about 1,600 packets/s. This difference reflects protocol-level overheads and event handling captured by OMNeT++ but abstracted in the analytical model. For consistency between throughput and utilization analysis, the effective controller service

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

TABLE I
NOTATION USED IN THE ASE-5G ANALYTICAL AND PERFORMANCE EVALUATION MODEL

Notation	Definition
λ_0	Initial external arrival rate of signaling messages entering the gNB.
λ_{bs}	Arrival rate at the gNB, including external arrivals and packets returning from the Software-Defined Networking (SDN) switch.
λ_{sw}	Arrival rate at the SDN switch, contributed by gNB-forwarded packets and packets looped back from the SDN controller.
λ_c	Arrival rate at the SDN controller, determined by the fraction of messages forwarded from the SDN switch.
q_{bs}	Forwarding probability from the gNB to the SDN switch.
q_{sw}	Forwarding probability from the SDN switch to the SDN controller.
q_c	Probability that packets processed at the SDN controller are returned to the SDN switch (loop-back probability).
μ_{bs}	Service rate of the gNB (requests/s).
μ_{sw}	Service rate of the SDN switch (requests/s).
μ_c	Service rate of the SDN controller (requests/s).
$E[D_{\text{TOT}}]$	Expected end-to-end delay of a registration request in the ASE-5G architecture.
$T_{\text{send},i}$	Transmission start time of the i-th request.
$T_{\text{receive},i}$	Reception completion time of the i-th request.
N	Total number of requests (observations) within the measurement interval.
$T_{\text{recovery},i}$	Recovery completion time of the i-th controller failure event.
$T_{\text{failure},i}$	Time at which the i-th controller failure is detected.
Latency	Time elapsed for a single request, calculated as $T_{\text{receive},i} - T_{\text{send},i}$.
Throughput	Number of packets successfully processed by the controller per second (packets/s).
Utilization	Percentage of controller workload relative to its maximum processing capacity.

TABLE II
QUANTITATIVE PERFORMANCE METRICS OF ASE-5G COMPARED WITH LITERATURE BASELINES

Metric	Standard 5G Core (3GPP UPF + GTP-U)	SoftNet	ASE-5G
Registration latency	Higher due to centralized anchoring and GTP-U overhead.	Not reported.	5–10 ms (low load), 10–25 ms (moderate load), >70 ms (saturation).
Intra-UPF / handover latency	Higher because of encapsulation and anchoring.	Not reported.	≈10 ms (low load), 10–20 ms (moderate load), sharp increase near saturation.
Inter-UPF handover latency	Higher due to UPF relocation and tunneling.	Not reported.	≈10–15 ms (low load), increases with load, >80 ms near saturation.
Controller throughput saturation	Limited by centralized processing.	Not reported.	≈1000 packets/s (OMNeT++), ≈1600 packets/s (queueing model).
Controller effective service rate	N/A	N/A	≈1000 packets/s.
SDN controller failover time	Implementation dependent.	Not reported.	5–10 s (low load), 15–25 s (moderate load), >80 s (saturation); inter-UPF: ≈20–30 s.
Controller utilization	Higher under centralized processing.	Not reported.	≈20–27% (moderate load, μ_{eff}-based).
GTP-U encapsulation	Yes	Not explicitly addressed.	No.
Mobility-aware flow redirection	Limited.	Partially supported.	Fully supported.
Multi-controller coordination	No.	Yes.	Yes (LC-PD aligned).
PFCP signaling	Standard PFCP.	Not the main focus.	Modified PFCP (deployment challenge acknowledged).

Note: The 3rd Generation Partnership Project (3GPP), Software-Defined Networking (SDN), GPRS Tunneling Protocol-User Plane (GTP-U), User Plane Function (UPF), Logically Centralized-Physically Distributed (LC-PD), Objective Modular Network Testbed in C++ (OMNeT++), and Packet Forwarding Control Protocol (PFCP).

rate is defined as $\mu_{eff} \approx 1,000$ packets/s. Using this effective rate, controller utilization remains relatively low ($\approx$ 20–27%) under moderate-load conditions, indicating that ASE-5G preserves processing headroom for mobility management and failover handling. The results further show that saturation occurs first at upstream elements, such as the gNB and OpenFlow switch, before the controller becomes a bottleneck. Failover performance analysis indicates that ASE-5G achieves bounded control-plane recovery times, with failover durations of 5–10 s under light load and 15–25 s under moderate load, increasing sharply only under extreme saturation.

Table II and Fig. 12 complement the numerical results by illustrating a capability-oriented architectural comparison across key dimensions, including control-plane adaptability, signaling efficiency, edge scalability, and failover resilience. The standard 5G core shows strong alignment with 3GPP SBA but is constrained by centralized control and GTP-U tunneling. SoftNet improves controller scalability and supports multi-controller coordination but lacks full SBA alignment and quantitative performance evaluation. ASE-5G achieves higher capability scores by eliminating GTP-U tunneling, enabling mobility-aware flow-level redirection, supporting LC-PD-aligned multi-controller coordination, and maintaining compatibility with 3GPP service-based procedures. While the use of modified

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,“Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture”, CommIT Journal 20(2), 263–282, 2026.

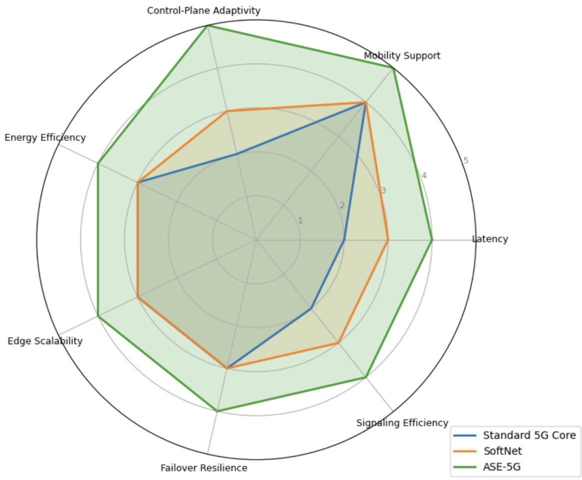

Fig. 12. Architecture comparison graph of ASE-5G, standard the 3rd Generation Partnership Project (3GPP) 5G core, and SoftNet across key control-plane and edge-network capabilities.

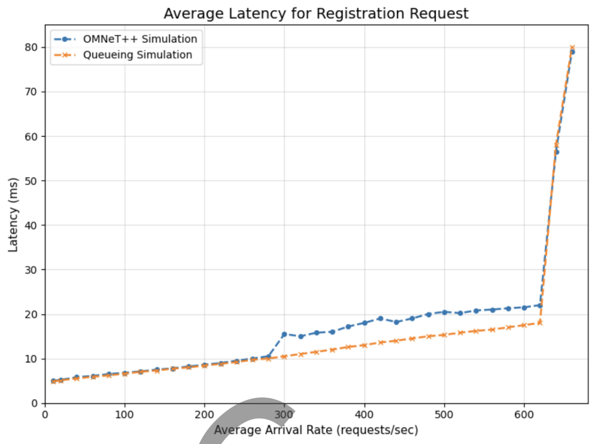

Fig. 13. Latency performance of registration request processing. Note: Objective Modular Network Testbed in C++ (OMNeT++).

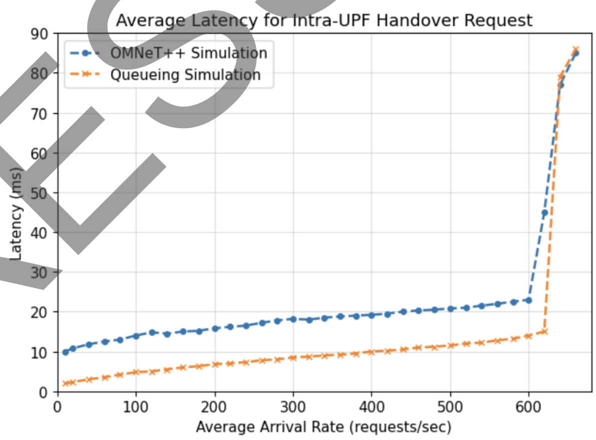

Fig. 14. Latency performance in intra-User Plane Function (UPF) handover procedures. Note: Objective Modular Network Testbed in C++ (OMNeT++).

PFCP signaling introduces deployment challenges, this limitation is explicitly acknowledged and motivates future standardization efforts [21].

Latency is defined as the time required for a request or packet to travel from the source to the destination across the network [32]. In this research, latency is calculated by subtracting the time when the request is received from the time when it is initially sent (see Eq. (9)). The final value of latency is obtained by taking the average across all requests observed during the experiment.

$$\text{Latency} = T_{\text{receive}} - T_{\text{send}},$$
$$\text{Average Latency} = \frac{1}{N}\sum_{i=1}^{N}\left(T_{\text{receive},i} - T_{\text{send},i}\right), \quad (9)$$

where $T_{\text{send},i}$ denotes the transmission start time of the i-th request, $T_{\text{receive},i}$ denotes the reception completion time of the i-th request, and N represents the total number of requests observed during the experiment.

The latency results for the registration request in ASE-5G are presented in Fig. 13. This trend reflects the onset of saturation, primarily due to the limited processing capability of the gNB, OpenFlow switch, and edge SDN controller. The observed bottleneck is aligned with previous research, where system delay increases rapidly once packet arrival rates approach the maximum processing capacity. The comparison between OMNeT++ and queueing simulations further validates the formulation in Eq. (9). Both approaches produce closely matching values, showing moderate latency growth under light to medium traffic and a sharp increase once saturation occurred. The consistency between analytical and simulation-based results reinforces the accuracy of the measurement method and the reliability of the experimental data.

The results for the Intra-Registration Area (Intra-RA) handover request are illustrated in Fig. 14. The simulation compares the average delay obtained from OMNeT++ and queueing-based analysis across varying arrival rates. At lower arrival rates, the observed delay remains relatively small, with OMNeT++ simulation results starting around 10 ms and queueing analysis yielding values below 5 ms. As the arrival rate increases, both results exhibit a gradual increase in delay. The OMNeT++ curve demonstrates a steady growth, maintaining values between 10 and 20 ms up to on the order of 600 requests/sec. In contrast, the queueing simulation consistently reports lower delays, showing the analytical model’s efficiency in estimating performance under light to moderate traffic conditions.

The results of the X_n-based handover with UPF

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,"Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture", CommIT Journal 20(2), 263–282, 2026.

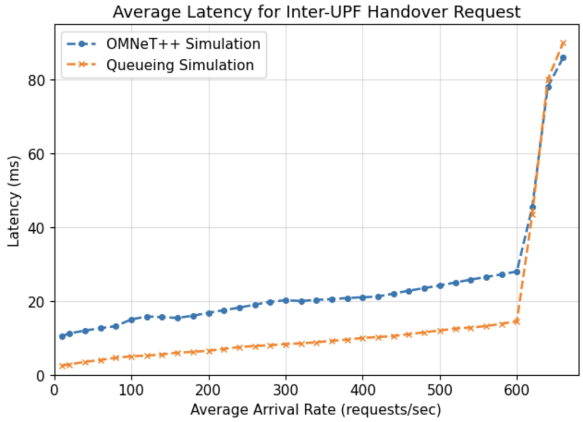

Fig. 15. Latency performance in inter-User Plane Function (UPF) handover procedures. Note: Objective Modular Network Testbed in C++ (OMNeT++).

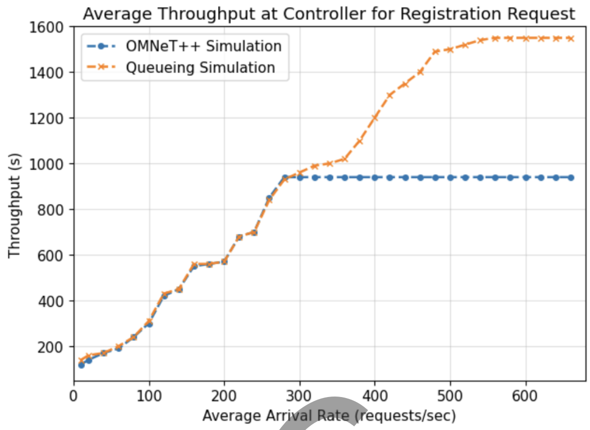

Fig. 16. Throughput performance for registration request process. Note: Objective Modular Network Testbed in C++ (OMNeT++).

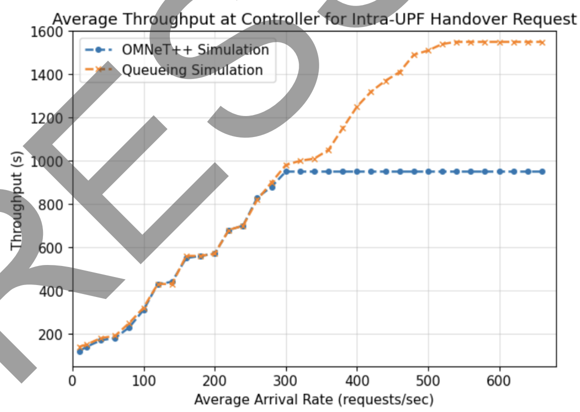

Fig. 17. Throughput performance in the intra-User Plane Function (UPF) handover request process. Note: Objective Modular Network Testbed in C++ (OMNeT++).

reallocation are presented in Fig. 15. A sharp increase in delay is observed when the system approaches saturation, with latency exceeding 80 ms at the maximum traffic load. This phenomenon aligns with queueing theory, where delays rise exponentially under heavy traffic conditions. These results confirm that the application of UPF reallocation in the handover procedure contributes to additional latency, and the discrepancy between the two approaches highlights the complexity of real network processing compared to abstract queueing models.

Next, throughput is defined as the rate at which the SDN controller successfully processes packets within a given unit of time [14]. As formulated in Eq. (10), throughput is computed by dividing the total number of packets processed by the duration of the simulation period. The formula is shown as follows:

$$\text{Throughput} = \frac{\text{Packets Processed}}{\text{Simulation Time}}. \quad (10)$$

This expression provides a direct measure of the controller's service efficiency. In the context of ASE-5G, throughput reflects the controller's ability to handle signaling packets generated during the registration and X_n-based handover procedures. The average throughput reported in this research is obtained by calculating the mean value across 30 independent simulation replications, ensuring that the results represent consistent and statistically reliable performance under different traffic loads. The average throughput observed at the SDN controller during the registration procedure is shown in Fig. 16.

On the other hand, the OMNeT++ simulation reaches saturation much earlier, stabilizing around 1,000 packets/sec once the arrival rate exceeds roughly 300 requests/sec. This earlier saturation occurs because OMNeT++ captures realistic packet-level behavior, including signaling exchanges, event scheduling, message handling, and protocol processing, which introduce additional control-plane overhead. Comparing the saturation points of the analytical model (≈ 1,600 packets/sec) and the OMNeT++ simulation (≈ 1,000 packets/sec) indicates that protocol-level and event-handling overhead account for approximately 37–40% reduction in effective throughput. This overhead arises from signaling exchanges, event scheduling, message parsing, and controller–switch interactions that are explicitly modeled in OMNeT++ but abstracted in the queueing formulation

The intra-UPF handover procedure is presented in Fig. 17. The results compare the OMNeT++ with the queueing simulation across a range of arrival rates. At lower arrival rates, both models exhibit a steady linear increase in throughput, indicating that the controller is able to efficiently process signaling packets without

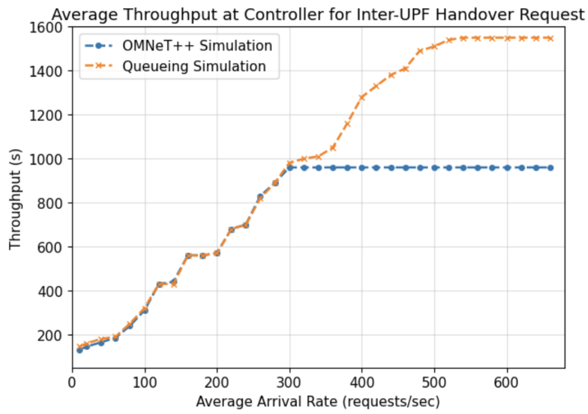

Fig. 18. Throughput performance in inter-User Plane Function (UPF) handover request process. Note: Objective Modular Network Testbed in C++ (OMNeT++).

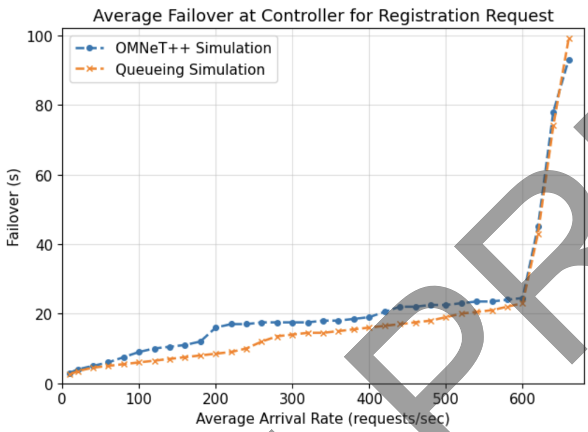

Fig. 19. Failover performance of registration request processing.

congestion. However, as the arrival rate continues to grow, the behaviors of the two models diverge. The queueing simulation reaches a saturation point at on the order of 1,600 packets/sec. It corresponds to the theoretical capacity derived from the base station processing time under idealized conditions. In contrast, the OMNeT++ simulation reaches a lower saturation point of around 1,000 packets/sec once the arrival rate exceeds roughly 300 requests/sec. This reduced saturation point reflects more realistic operating conditions, as OMNeT++ accounts for packet-level overheads, such as protocol processing, signaling interactions, event scheduling, and system delays. The inter-UPF handover requests are shown in Fig. 18. These saturation points represent extreme operating conditions, whereas the utilization analysis focuses on moderate load regimes that reflect practical deployment scenarios.

Next, to address the apparent inconsistency between throughput saturation and SDN controller utilization, it is important to distinguish between the theoretical service capacity derived from the controller processing time and the effective throughput observed in OMNeT++, which includes protocol and event-handling overhead. Based on the configured processing time of 0.2 ms/request, the theoretical maximum service rate is 1/0.0002 = 5,000 requests/s. However, OMNeT++ captures additional operations (e.g., event scheduling, message handling, and protocol processing), which reduce the effective throughput and lead to earlier saturation around 1,000 packets/s.

Failover refers to the process of transferring network control from the primary SDN controller to a backup controller when the primary controller becomes unavailable [21]. This mechanism is essential for maintaining network continuity and preventing service disruption in the proposed ASE-5G architecture. In this research, the duration of failover is measured by calculating the time difference between the moment when the failure is detected and the moment when the backup controller fully restores network operations. The failover time for each event is computed using Eq. (11), where the recovery time T_{recovery} is subtracted from the failure detection time T_{failure}

$$\text{Failover Time} = T_{\text{recovery}} - T_{\text{failure}},$$

$$\text{Average Failover Time} = \frac{1}{N}\sum_{i=1}^{N}(T_{\text{recovery},i} - T_{\text{failure},i}). \quad (11)$$

To obtain a reliable and representative metric, the average failover time is calculated by taking the mean value across all observed failover events, as defined in Eq. (11). In this formulation, $T_{\text{failure},i}$ indicates the time at which the ith failure event occurs, $T_{\text{recovery},i}$ represents the time at which the system successfully completes the recovery process for the corresponding event, and N denotes the total number of failure-recovery cycles recorded during the simulations. By averaging the failover time across multiple events, the resulting metric provides a reliable and statistically representative measure of the network recovery performance under varying network conditions.

The results of the failover analysis for the registration request are presented in Fig. 19. At lower arrival rates below 200 requests/s, the average failover time remains relatively short, ranging from 5 to 10 s in OMNeT++ and slightly lower in the queueing-based model. Under light traffic conditions, the system can rapidly switch to the backup controller without introducing significant service disruption. As the arrival rate increases to the range of 200 to 600 requests/s, the average failover time gradually increases, reaching

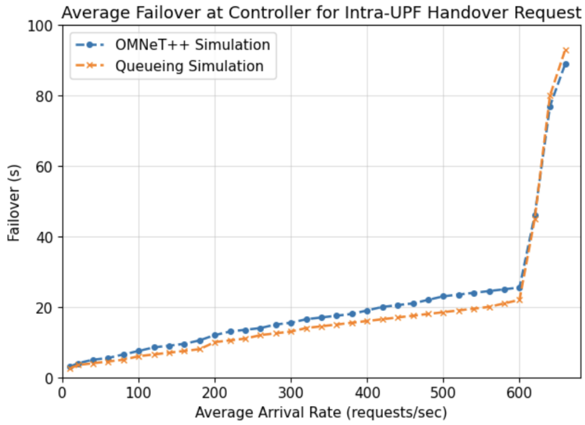

Fig. 20. Failover performance in intra-User Plane Function (UPF) handover procedures. Note: Objective Modular Network Testbed in C++ (OMNeT++).

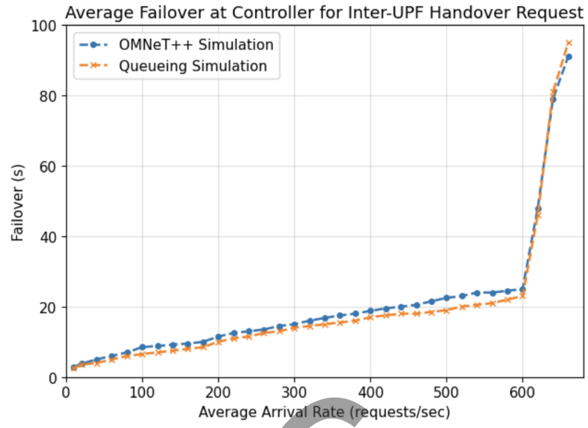

Fig. 21. Failover performance in inter-User Plane Function (UPF) handover procedures. Note: Objective Modular Network Testbed in C++ (OMNeT++).

approximately 15 to 25 s. This trend is consistent with queueing theory, in which higher arrival rates increase the controller workload, thereby extending the recovery process. A critical escalation is observed once the arrival rate exceeds 5,000 requests/s. In this region, Eq. (11) produces average failover times exceeding 80 s for both the OMNeT++ simulation and the queueing-based model, indicating the onset of system saturation. The OMNeT++ simulation consistently reports slightly higher failover times because it incorporates additional packet-level overheads, including signaling, synchronization, and protocol interactions, which are not considered in the analytical queueing formulation.

The failover performance during the intra-UPF handover procedure is presented in Fig. 20. When the arrival rate exceeds 5000 requests/sec, Eqs. (11) results in a sharp escalation of failover time, with both approaches reporting values beyond 80 s. The OMNeT++ simulation produces slightly higher results than the analytical model, which can be attributed to the inclusion of detailed network overhead such as packet signaling and synchronization delays.

The failover performance during the inter-UPF handover procedure is illustrated in Fig. 21. According to Eq. (11), the failover time is defined as the interval between the detection of a controller failure and the successful restoration of network operations by the backup controller. At low arrival rates (below 200 requests/s), both the OMNeT++ simulation and the analytical queueing model report relatively short failover durations, typically ranging from 5 to 10 s. This behavior indicates that, under light traffic conditions, the proposed ASE-5G architecture can recover from controller failures with minimal service disruption.

As the arrival rate increases to between 200 and 600 requests/s, the average failover time gradually increases to approximately 20 to 30 s because of the higher controller workload and increased queueing delays. Compared with the registration and intra-UPF handover scenarios, the difference between the OMNeT++ simulation and the analytical queueing model becomes more noticeable. The OMNeT++ simulation consistently reports slightly higher failover times because it explicitly models protocol-level interactions, packet processing, event scheduling, controller synchronization, and signaling overhead, which are abstracted in the analytical queueing model.

A substantial increase in failover time is observed once the arrival rate exceeds approximately 5,000 requests/s. Under this traffic condition, both the OMNeT++ simulation and the analytical queueing model predict failover times exceeding 80 s, indicating that the system has entered the saturation region. The larger discrepancy between the two approaches reflects the additional processing overhead introduced by UPF relocation during inter-UPF handover, which requires more complex control-plane operations than the intra-UPF handover procedure.

It should be emphasized that the failover durations reported in this research, which exceed 80 s under saturation conditions, correspond exclusively to control-plane recovery processes rather than user-plane packet delivery. The latency requirements defined for 5G URLLC apply to user-plane packet transmission and require end-to-end delays on the order of milliseconds, whereas controller failover represents an infrequent network-management operation. In practical deployments, controller failover is typically mitigated through controller redundancy, state synchronization, and pre-installed backup forwarding rules to minimize service

Cite this article as: V. Monita, N. H. Lutfianto, B. Prasetya, and A. S. Karnyoto,"Performance evaluation of control plane signaling in 5G: A novel adaptive software-edge architecture", CommIT Journal 20(2), 263–282, 2026.

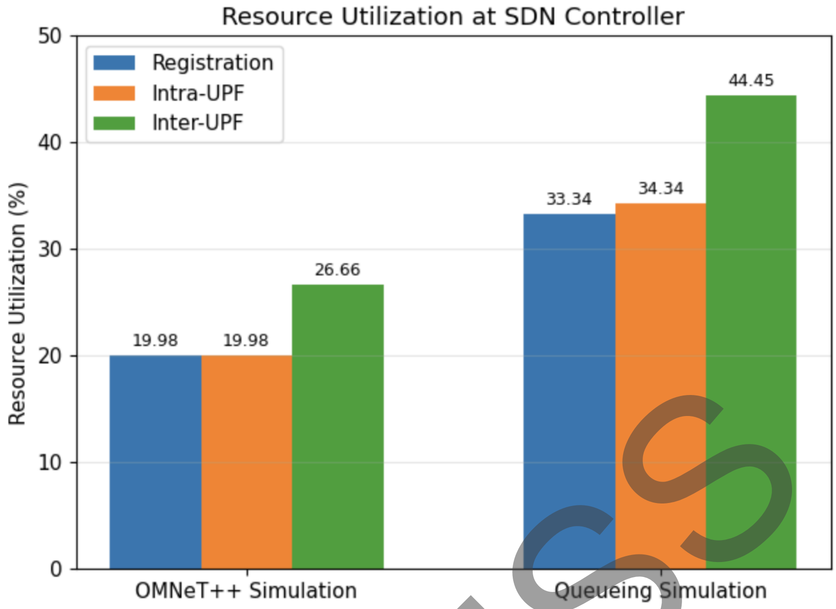

Fig. 22. SDN controller utilization under different traffic load conditions, computed using the effective service rate derived from Objective Modular Network Testbed in C++ (OMNeT++) simulations. Note: User Plane Function (UPF).

interruption. Therefore, the failover results presented in this research represent worst-case recovery scenarios under saturated traffic conditions and should not be interpreted as the user-plane latency experienced by URLLC services.

Next, resource utilization represents the proportion of the SDN controller's available processing capacity that is actively used to process signaling requests. It reflects the operational workload experienced by the controller during network operation. As expressed in Eq. (12), controller utilization is calculated as the ratio between the observed controller throughput and the effective controller service capacity,

$$\text{Utilization} = \frac{TP_{\text{controller}}}{C_{\text{controller}}} \times 100\%, \tag{12}$$

where $TP_{\text{controller}}$ denotes the throughput measured at the SDN controller, and $C_{\text{controller}}$ represents the effective processing capacity of the controller.

Based on the configured processing time of 0.2 request/ms, the theoretical maximum service capacity of the SDN controller is 5,000 requests/s. However, the OMNeT++ simulation explicitly models additional control-plane overheads, including protocol processing, signaling exchanges, event scheduling, queue management, and controller-switch interactions. These overheads reduce the effective processing capacity, resulting in an observed throughput saturation of approximately 1,000 packets/s. Therefore, to ensure consistency between the analytical evaluation and the simulation results, controller utilization is calculated with respect to the effective service capacity obtained from the OMNeT++ simulation. The controller utilization under different traffic loads is illustrated in Fig. 22.

IV. Conclusion

The research proposes and evaluates the ASE-5G architecture, which integrates SDN programmability with edge computing intelligence to improve control-plane signaling efficiency and resilience in 5G core networks. Through OMNeT++ simulations and analytical modeling, the results demonstrate that ASE-5G achieves lower latency, stable throughput, controlled resource utilization, and improved robustness under controller failover in light to moderate traffic conditions. These findings confirm the effectiveness of combining SDN-based forwarding with edge-assisted control-plane coordination for dynamic signaling management. Performance saturation appears under extreme signaling loads. Therefore, ASE-5G is best positioned for edge-centric/private 5G and enterprise deployments rather than carrier-scale cores. Within this scope, ASE-5G simplifies mobility handling and avoids GTP-U tunneling while keeping SDN controller headroom for control tasks. However, the key deployment barriers remain regarding the use of modified PFCP and the lack of a standardized east–west interface for multi-controller interoperability.

Future work should prioritize the formal specification and standardization of eas-west controller communication, including interoperable signaling mechanisms and control abstractions, potentially aligned with ongoing 3GPP and European Telecommunications Standards Institute (ETSI) initiatives. Further research may also explore scalable load distribution mechanisms and selective integration of learning-based techniques to enhance predictive adaptation. Finally, validation on realistic testbeds and controlled experimental platforms will be essential to assess the practical feasibility of ASE-5G beyond simulation-based evaluation, particularly once interface standardization constraints are addressed.

Acknowledgement

This research was supported by the Internal Research Program of Telkom University, Bandung, Indonesia, under Grant No. RCMS/RES/Penelitian Internal/15608/1/2025. The authors also gratefully acknowledge Bina Nusantara University, Indonesia, for its academic support.

Author Contribution

Designed the ASE-5G architecture and supervised the research, V. M.; Prepared and performed the analytical modeling and interpretation of results, V. M.; Contributed to the design of the SDN-based data plane and queueing model formulation, V. M.; Designed and conducted OMNeT++ simulation scenarios, N. H. L.; Ran the simulations and collected performance data, N. H. L.; Assisted in the analysis of simulation and analytical results, B. P.; Contributed to the literature review, discussion of SDN and edge computing integration, and manuscript refinement, A. S. K.

Data Availability

The generated datasets are available from the corresponding author upon reasonable request, Vivi Monita. The data supporting the findings of this research are obtained from OMNeT++ simulations and analytical queueing models developed by the authors. All simulation settings, parameters, and evaluation procedures are fully described in the manuscript to ensure reproducibility.

References

[1] Z. Mlika and S. Cherkaoui, "Network slicing with MEC and deep reinforcement learning for the Internet of vehicles," *IEEE Network*, vol. 35, no. 3, pp. 132–138, 2021.

[2] Y. Liu, M. Peng, G. Shou, Y. Chen, and S. Chen, "Toward edge intelligence: Multiaccess edge computing for 5G and Internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6722–6747, 2020.

[3] A. Perdigão, J. Quevedo, and R. L. Aguiar, "Automating 5G network slice management for industrial applications," *Computer Communications*, vol. 229, pp. 1–19, 2025.

[4] A. Nassar and Y. Yilmaz, "Deep reinforcement learning for adaptive network slicing in 5G for intelligent vehicular systems and smart cities," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 222–235, 2021.

[5] S. Wijethilaka and M. Liyanage, "Survey on network slicing for Internet of things realization in 5G networks," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 957–994, 2021.

[6] Ericsson, "Ericsson mobility report," 2026. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report

[7] V. Monita, R. Munadi, and I. D. Irawati, "Video streaming service performance with high availability using SDN protocols," *ICIC Express Letters*, vol. 18, no. 9, pp. 895–905, 2024.

[8] A. Domeke, B. Cimoli, and I. T. Monroy, "Integration of network slicing and machine learning into edge networks for low-latency services in 5G and beyond systems," *Applied Sciences*, vol. 12, no. 13, pp. 1–22, 2022.

[9] H. Wang, S. Chen, H. Xu, M. Ai, and Y. Shi, "SoftNet: A software defined decentralized mobile network architecture toward 5G," *IEEE Network*, vol. 29, no. 2, pp. 16–22, 2015.

[10] L. Albshaier, S. Almarri, and A. Albuali, "Federated learning for cloud and edge security: A systematic review of challenges and AI opportunities," *Electronics*, vol. 14, no. 5, pp. 1–60, 2025.

[11] M. K. Benbraika, O. Kraa, Y. Himeur, K. Telli, S. Atalla, and W. Mansoor, "Enhancing 5G vehicular edge computing efficiency with the Hungarian algorithm for optimal task offloading," *Computers*, vol. 13, no. 11, pp. 1–15, 2024.

[12] P. Vidhya, K. Subashini, R. Sathishkannan, and S. Gayathri, "Dynamic network slicing based resource management and service aware Virtual Network Function (VNF) migration in 5G networks," *Computer Networks*, vol. 259, 2025.

[13] W. K. Chiang and Y. H. Shang, "ES-5G: A novel edge-based SDN-enabled 5G architecture for lower latency," in *Proceedings of the 2023 6th International Conference on Information Science and Systems*. Edinburgh, United Kingdom: Association for Computing Machinery, Aug. 11–13, 2023, pp. 143–153.

[14] A. C. Baktır, A. Özgövde, and C. Ersoy, "End-

to-end network slicing for edge computing optimization," *Future Generation Computer Systems*, vol. 157, pp. 516–528, 2024.

[15] Q. V. Pham *et al.*, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE Access*, vol. 8, pp. 116 974–117 017, 2020.

[16] A. Bazzi, R. Bomfin, M. Mezzavilla, S. Rangan, T. Rappaport, and M. Chafii, "Upper mid-band spectrum for 6G: Vision, opportunity and challenges," *IEEE Communications Magazine*, vol. 64, no. 1, pp. 206–212, 2026.

[17] S. Ehsanfar, A. Bazzi, K. Mößner, and M. Chafii, "Hypothesis testing on FMCW and OFDM for Joint communication and radar in IEEE 802.11 bd," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*. Rome, Italy: IEEE, May 28–June 1, 2023, pp. 464–469.

[18] A. Bazzi, M. Ying, O. Kanhere, T. S. Rappaport, and M. Chafii, "ISAC imaging by channel state information using ray tracing for next generation 6G," *IEEE Journal of Selected Topics in Electromagnetics, Antennas and Propagation*, vol. 1, no. 1, pp. 135–149, 2025.

[19] A. Basta, W. Kellerer, M. Hoffmann, H. J. Morper, and K. Hoffmann, "Applying NFV and SDN to LTE mobile core gateways, the functions placement problem," in *Proceedings of the 4^{th} Workshop on All Things Cellular: Operations, Applications, & Challenges*. Illinois, USA: Association for Computing Machinery, Aug. 22, 2014, pp. 33–38.

[20] L. R. Nair, P. Mukherjee, and S. Chatterjee, "Securing tomorrow's connectivity: Developing defenses for 5G-enhanced IoT technologies," in *Securing the Connected World: Exploring Emerging Threats and Innovative Solutions*. Springer, 2025, pp. 263–295.

[21] P. Fondo-Ferreiro, F. Gil-Castineira, F. J. González-Castaño, and D. Candal-Ventureira, "A software-defined networking solution for interconnecting network functions in service-based architectures," *IEEE Access*, vol. 10, pp. 19 905–19 916, 2022.

[22] S. Moh, "Resource allocation techniques in aerial-assisted vehicular edge computing: A review of recent progress." *Electronics*, vol. 14, no. 18, 2025.

[23] P. Garcia Lopez *et al.*, "Edge-centric computing: Vision and challenges," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 5, pp. 37–42, 2015.

[24] S. Bolettieri, D. T. Bui, and R. Bruno, "Towards end-to-end application slicing in multi-access edge computing systems: Architecture discussion and proof-of-concept," *Future Generation Computer Systems*, vol. 136, pp. 110–127, 2022.

[25] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[26] B. U. Kazi, M. K. Islam, M. M. H. Siddiqui, and M. Jaseemuddin, "A survey on software defined network-enabled edge cloud networks: Challenges and future research directions," *Network*, vol. 5, no. 2, pp. 1–28, 2025.

[27] B. J. Chang and Y. T. Lin, "Optimizing differentiated SFCs of slicing using virtual/physical network resources for SDN-based flowing and steering in B5G/5G networking," *Computer Networks*, vol. 259, 2025.

[28] S. R. Aldaeabool, W. H. Al-Zubaedi, M. F. Abbod, and H. S. Al-Raweshidy, "Signaling load reduction in 5G network based using Open-RAN architecture and SDN," in *2025 4^{th} International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. Dhaka, Bangladesh: IEEE, Jan. 11–12, 2025, pp. 199–204.

[29] R. Bruschi, J. F. Pajo, F. Davoli, and C. Lombardo, "Managing 5G network slicing and edge computing with the MATILDA telecom layer platform," *Computer Networks*, vol. 194, 2021.

[30] K. S. Sahoo, S. Mohanty, M. Tiwary, B. K. Mishra, and B. Sahoo, "A comprehensive tutorial on software defined network: The driving force for the future internet technology," in *Proceedings of the International Conference on Advances in Information Communication Technology & Computing*. Bikaner, India: Association for Computing Machinery, Aug. 12–13, 2016, pp. 1–6.

[31] J. Chen, X. Liang, J. Xue, Y. Sun, H. Zhou, and X. Shen, "Evolution of RAN architectures toward 6G: Motivation, development, and enabling technologies," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 1950–1988, 2024.

[32] D. M. Gonçalves, L. F. Bittencourt, and E. R. M. Madeira, "Overhead and performance of dynamic network slice allocation for mobile users," *Future Generation Computer Systems*, vol. 160, pp. 739–751, 2024.

[33] M. J. Kulkarni and M. Ramamoorthy, "Resource allocation and secure channel model for service-

dependent mobile edge computing," in *2024 9th International Conference on Communication and Electronics Systems (ICCES)*. Coimbatore, India: IEEE, Dec. 16–18, 2024, pp. 958–964.

[34] Y. Zhang, K. Zhao, Y. Yang, and Z. Zhou, "Real-time service migration in edge networks: A survey," *Journal of Sensor and Actuator Networks*, vol. 14, no. 4, pp. 1–40, 2025.

[35] D. Xiao, S. Chen, W. Ni, J. Zhang, A. Zhang, and R. Liu, "A sub-action aided deep reinforcement learning framework for latency-sensitive network slicing," *Computer Networks*, vol. 217, 2022.

[36] K. Hejja, S. Berri, and H. Labiod, "Network slicing with load-balancing for task offloading in vehicular edge computing," *Vehicular Communications*, vol. 34, 2022.

[37] A. J. Zharabad, S. Yousefi, and T. Kunz, "Network slicing in virtualized 5G Core with VNF sharing," *Journal of Network and Computer Applications*, vol. 215, 2023.

[38] F. Azeez, "SDN-based edge computing," in *International Conference on Mathematical Modeling and Computational Science*. Springer, 2025, pp. 86–95.

[39] S. F. Ismail, "Real-time SDN—IoT integrated framework for intelligent emergency vehicle prioritization in smart cities," *Mesopotamian Journal of Computer Science*, vol. 2025, pp. 186–201, 2025.

[40] V. Dhamini, Y. Subathra, V. Aruna, J. S. B. Jothi, A. Sathish, and G. Ali, "Scalable multi-connectivity approaches for AR/VR traffic management in 5G networks," in *2025 7th International Conference on Inventive Material Science and Applications (ICIMA)*. Namakkal, India: IEEE, May 28–30, 2025, pp. 1415–1420.

[41] S. Bhaskaran and S. Muthuraman, "A comprehensive study of resource provisioning and optimization in edge computing," *Computers, Materials & Continua*, vol. 83, no. 3, pp. 5037–5070, 2025.

[42] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 196–248, 2019.

[43] B. Liang, M. A. Gregory, and S. Li, "Multi-access edge computing fundamentals, services, enablers and challenges: A complete survey," *Journal of Network and Computer Applications*, vol. 199, 2022.

[44] X. Tong, X. Li, and Y. Liu, "An auction-based distributed network slicing scheme for resource allocation in satellite-UAV integrated networks," *Computer Communications*, vol. 210, pp. 58–68, 2023.

[45] M. Barbosa, M. Silva, E. Cavalcanti, and K. Dias, "Open-source 5G core platforms: A low-cost solution and performance evaluation," in *2025 International Conference on Information Networking (ICOIN)*. Chiang Mai, Thailand: IEEE, Jan. 15–17, 2025, pp. 99–104.

[46] X. Lin, "3GPP evolution from 5G to 6G: A 10-year retrospective," *Telecom*, vol. 6, no. 2, pp. 1–30, 2025.