# ES-5G: A Novel Edge-based SDN-enabled 5G Architecture for Lower Latency

Wei-Kuo Chiang*
Department of Computer Science and Information Engineering, National Chung Cheng University
wkchiang@cs.ccu.edu.tw

Yi-Hsin Shang
Department of Computer Science and Information Engineering, National Chung Cheng University
a0933520516@gmail.com

## ABSTRACT

With the fast development of the Internet and the increasing data traffic forwarded in the wireless network, the mobile network raises the requirement of bandwidth and latency. To meet this trend and provide better user Quality of Service (QoS), we deploy a popular paradigm called Software Defined Networking (SDN) with the existing 5G network and combine it with the concept of edge computing to achieve the purpose of higher data rate and lower latency. First, we replace the data plane in the 5G network with the SDN network and deploy the SDN network in the edge network close to the user. This new architecture is called the Edge-based SDN-enabled 5G network (ES-5G). It can shorten the data forwarding path, reduce the data process time, and reduce the burden on the backhaul network that forwards the data to the core network. After that, we modify the existing basic procedures to satisfy our design. We propose a new mechanism to pre-establish the forwarding path and design the new handover procedure. This mechanism is less processed when the user moves to other gNodeB (gNB) service domains. Then, we calculate the end-to-end latency from forwarding the packets and the message exchanges in the handover procedures to compare with the original 5G network. Finally, we verify that our performance analysis is correct via OMNeT++.

## CCS CONCEPTS

• **Networks** → Network architectures; Network design principles; Layering.

## KEYWORDS

5G, Software Defined Network (SDN), edge computing, handover procedure, OMNeT++

**ACM Reference Format:**
Wei-Kuo Chiang and Yi-Hsin Shang. 2023. ES-5G: A Novel Edge-based SDN-enabled 5G Architecture for Lower Latency. In *2023 The 6th International Conference on Information Science and Systems (ICISS 2023), August 11–13, 2023, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3625156.3625178

*Corresponding author



*ICISS 2023, August 11–13, 2023, Edinburgh, United Kingdom*

ACM ISBN 979-8-4007-0820-6/23/08.
https://doi.org/10.1145/3625156.3625178

## 1 INTRODUCTION

The fast development of the Internet, mobile devices, and the Internet of Things (IoT) and the increasing data traffic forwarded in the wireless network. The 5G network has been defined to satisfy future development that has better performance in data rate and latency [1].

Three typical service scenarios in the 5G network are considered enhanced Mobile Broadband (eMBB), ultra-Reliable and Low Latency Communications (uRLLC), and massive Machine Type Communications (mMTC). The eMBB focuses on services with high data rates, such as virtual reality (VR), augmented reality (AR), etc. The uRLLC provides latency-sensitive and highly reliable services (e.g., self-driving). Finally, the mMTC focuses on services with massive connections with IoT devices.

In the eMBB and uRLLC service scenarios, we consider reducing data latency a critical challenge. In [2], A. Huang et al. propose an SDN-based MEC framework and implement it to prove latency reduction and traffic offloading; the complex concept of MEC can be referred to the MEC white paper [3]. Softbox [4] is another new architecture combining SDN, NFV, and MEC technologies. M. Moradi virtualized and re-architectured the EPC functions for the customized UE container and deployed the UE containers in the edge network to reduce the latency from forwarding the packets to the core network. In addition, the authors also deliver packets via the SDN switches to avoid the latency of the GPRS Tunneling Protocol (GTP). R. Trivisonno et al. in [5] present a novel SDN-based 5G network architecture and virtualize the control plane functions to re-configure them in three types of SDN Controllers. This architecture has better performance in default bearer establishment and user-plane latency. Moreover, this paper also describes basic procedures. A. Jain et al. in [6] describe and evaluate the design of two types of LTE EPC networks, one based on SDN principles and the other based on network function virtualization (NFV). The result shows that NFV-based EPC is better suited for networks with high signaling traffic because the SDN Controller quickly becomes the bottleneck in the SDN-based EPC. On the other hand, the SDN-based EPC is better suited for networks with high data plane traffic because the SDN switches are more optimized for packet forwarding than virtualized software appliances. M. T. Raza et al. in [7] propose refactoring IMS NFs modules and a "prefetching algorithm." The Media Resource Function Processor (MRFP) can predict future packets and pre-fetch control instructions via the algorithm and pipeline data packets processing and fetch its control instructions to reduce the media latency in the IMS system. Takamasa Ochiai et al. propose an approach that can reduce latency and bring higher throughput called the Moving Cell Support Protocol based on a Locator/ID split approach (MocLis) [8]. This approach modifies the IPv6 prefix to forward the package between eNodeB and UPF without using the GTP tunneling.

UE mobility changes the connection between the user devices and the gNBs. Therefore, mobility management also is a significant challenge in the 5G network. H. Ko et al. in [9] propose an SDN approach for Distributed Mobility Management (DMM). It implements the SDN Controller's location and handover management functions and distributes the packet forwarding function at access routers (AR). This approach can achieve the optimal forwarding path from a correspondence node (CN) to the new AR without any tunneling overhead. To address the limitations of centralized mobility management in cellular networks, H. Jin et al. in [10] present the virtualized mobility management approach. The authors divide the MME into three virtual components, including Mobility Signaling Forwarding (MSF), Mobility Management Processor (MMP), and Data Query Update (DQU), and propose the algorithm to find out the optimal component placement. The mobility management procedures, including the handover procedures, how to reduce the latency in handover procedures, and session continuity, are also significant challenges in the 5G network. S. Kuklinski et al. in [11] discuss three types of SDN technology applied in the mobile network: centralized SDN, semi-centralized SDN, and hierarchical SDN. They describe their pros and cons and the problem of handover management. In [12], M. Erel-Özçevik and B. Canberk remove the handover execution phase in the Xn-based Handover procedure because the Software-Defined Ultra-Dense Network (SDUN) Controller can pre-determine and pre-synchronize T-eNodeB to reduce the end-to-end latency.

## 1.1 Motivation

In [4, 6, 9] all utilize the SDN technology to reduce the forwarding packets latency, and their deployment of the SDN Controller is centralized. However, it has a big problem that the SDN Controller may overload and cause a bottleneck when its service range is a global network. The 5G network is an extensive scale network, so the centralized SDN Controller is unsuitable for the 5G network. A. Gasmelseed et al. in [13] and C. N. Tadros et al. in [14] prove that the distributed SDN Controller performs better than the centralized SDN Controller in latency. A. Gasmelseed et al. in [13] compare the performance of centralized and distributed SDN Controller using load balance application with three algorithms: the Round Robin algorithm, Weighted Round Robin algorithm, and Random algorithm. The results show that the distributed SDN Controller performs better in response time and the number of transactions than the centralized SDN Controller. C. N. Tadros et al. in [14] compare the performance of three different deployment types of the SDN Controller via generating the data traffic (e.g., HTTP traffic) to calculate their latency; the detailed three different deployment types are in section 2.2. More detail about the distributed SDN Controller can refer to [15].

The SDN is a flow-based technique whose performance is based on the flow-setup time and the number of flows the SDN Controller can handle [16]. There are two methods to set up the flow tables: proactive and reactive. The proactive method is to pre-set up the flow tables before the packets arrive at the OpenFlow Switch. This method causes negligible latency because the OpenFlow Switch knows how to handle the packets. The other method is that after the packets have arrived at the OpenFlow Switch, the OpenFlow Switch needs to forward the packets to the SDN Controller to handle them. This method causes more latency of the SDN Controller processing time and updating the flow table in the OpenFlow Switch. We modify the procedures in our ES-5G network because this paper provides the concept that the proactive method can bring lower latency.

## 1.2 Contribution

- In the paper, we refer to 5G technical specifications [17, 18], the SDN architecture specification [19], and the OpenFlow switch specification [20] to enable the SDN technology in the original data plane of the 5G network and combine the concept of the edge computing to propose a new architecture is called ES-5G.
- We re-designed the basic procedures and created the new handover procedures. We decided to offer the information on the UE Registration Area to the SDN Controller so that the SDN Controller can pre-establish the forwarding paths in the PDU Session Establishment procedure. This mechanism can bring a faster handover procedure when the UE mobility in the Registration Area.
- We analyze and compare the end-to-end latency performance of our design and the original 5G reference point architecture. Finally, we simulate two architectures to prove our performance analysis is correct.

The rest of this paper is organized as follows. Section 2 overviews the related work on the 5G network and SDN. Section 3 presents our new architecture called the ES-5G network and its component functions and procedures—the results of the performance evaluation in Section 4. In Section 5, we simulate the previous review to verify our analysis using the OMNeT++ software. Finally, we conclude the thesis and its future work with Section 6.

# 2 RELATED WORKS

The fast development of communication technology and the increasing number of IoT devices need to connect to the Internet. The European Telecommunications Standards Institute (ETSI) has defined the fifth Generation Mobile Networks (5G) to support future development. The 5G network can bring higher data rates and lower latency than old-generation mobile networks. SDN technology has been defined by the Open Networking Foundation (ONF) for improving existing network architecture. Its key concept is decoupling the control and data planes, allowing the handling the traffic network via software. The SDN Controller controls and manages the underlying network devices, and the network devices only forward the packets. These fundamental tenets offer flexible and straightforward management/configuration in the SDN network. This section will describe the systems and technologies related to this paper.

## 2.1 5G Network Architecture

The 5G network comprises an access and core network, as Figure 1 shows. The 5G architecture is defined by the 3rd Generation Partnership Project(3GPP); it has two ways to represent the interface

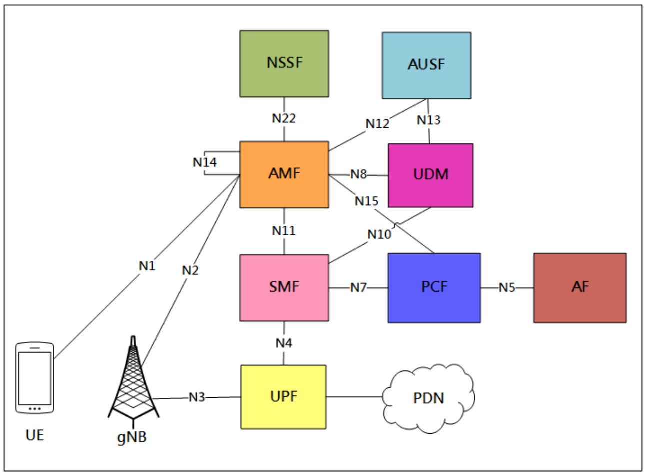


**Figure 1: 5G network architecture**

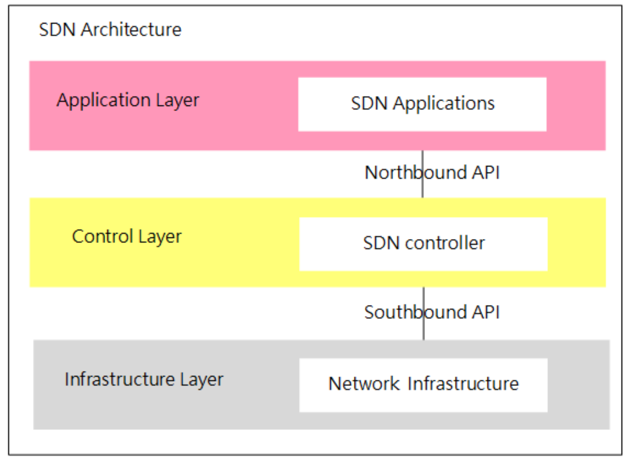


**Figure 2: SDN network architecture**

between network functions. One way is the service-based representation; the other is the reference point representation. The paper describes the 5G architecture in reference point representation. The 5G architecture in service-based representation can refer to [17].

The access network consists of multiple gNBs, and their primary function is to help users connect to the 5G network.

The core network consists of multiple components; the following will describe their primary functions. However, suppose the components' functions are related to network slicing technology. In that case, we will not discuss them, like Network Slice Selection Function (NSSF) and Application Function (AF).

- The Access and Mobility Management Function (AMF) is responsible for registration, mobility, connection, and access authentication/authorization.
- The Session Management Function (SMF) is responsible for session management, UE IP address allocation and management, and User Plane Function (UPF) traffic steering configuration.
- The UPF handles packet routing, forwarding, inspection, and data plane QoS.
- The Authentication Server Function (AUSF) is an authentication server.
- The Unified Data Management (UDM) is responsible for subscription management and user identification handling (e.g., Subscription Permanent Identifier (SUPI)).
- The Policy Control Function (PCF) provides policy rules and enforces them.

## 2.2 SDN Network Architecture

Figure 2 shows the SDN network architecture. It is composed of three layers: the application layer, the control layer, and the infrastructure layer. The infrastructure layer includes essential network devices, like switches or routers. The control layer consists of a logically centralized SDN Controller that controls the underlying infrastructure via the Southbound Applications Programming Interface (S-API) and communicates the application layer via the Northbound API (N-API). Nowadays, the common N-API is a Representational State Transfer (REST) API; the common S-API is OpenFlow.

## 2.3 SDN Controller Deployment

The SDN Controller deployment will affect network performance, such as latency and connectivity. Two common types of SDN control plane architectures are centralized and distributed architectures.

Figure 3 presents a new way to deploy an SDN Controller called Logically Centralized-Physically Distributed (LC-PD) [14]. It resembles the multi-core SDN Controller to manage the whole network. However, every SDN Controller only manages its domain's underlying resources (e.g., OpenFlow Switches) and shares their underlying network information via network synchronization. Finally, the authors compare the previous three types of latency and throughput performance. The result shows that the LC-PD architecture is the best one.

The LC-PD architecture is better than a distributed architecture. Each OpenFlow Switch is connected to only one SDN Controller in LC-PD architecture. Therefore, only one SDN Controller processes the OpenFlow Switch and shares its caching information with other SDN Controllers. However, in a distributed architecture, all SDN Controllers may process the same OpenFlow Switch and synchronize the exact information of the OpenFlow Switch.

## 2.4 OpenFlow Switch Architecture

The SDN Controller provides a programmable platform so that the administrator can add, modify, and delete flow tables in the OpenFlow Switches via the OpenFlow protocol.

The OpenFlow Switch has five main parts: the protocol, channel, group table, meter table, and flow table, as shown in Figure 4.

The OpenFlow protocol is between each SDN Controller and each OpenFlow Switch.

The OpenFlow channel is a secure channel that connects each OpenFlow Switch to an SDN Controller. It is usually encrypted using TLS but may run directly over TCP.

- The group table represents more complex forwarding methods like multipath or broadcast.

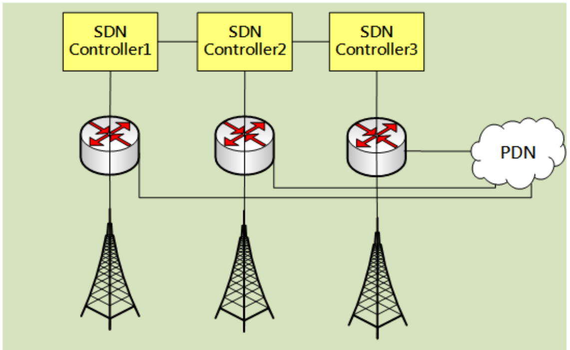


**Figure 3: Logically Centralized-Physically Distributed (LC-PD) control plane architecture**

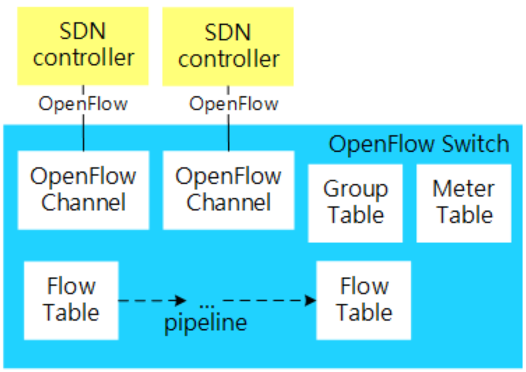


**Figure 4: OpenFlow Switch architecture**

| | | | | | | |
|---|---|---|---|---|---|---|
| Match Field | Priority | Counters | Instructions | Timeouts | Cookie | Flags |

**Figure 5: Main components of the flow entry**

- The meter table implements rate limiting and QoS operation to the packets.
- The flow table consists of flow entries that do not exist by default. Figure 5 shows the flow entry's main components.

## 3 EDGE-BASED SDN-ENABLED 5G NETWORK

This section introduces the proposed ES-5G network architecture, functions, modified basic procedures, and new procedures of the ES-5G network.

### 3.1 ES-5G Network Architecture

This section discusses deploying the SDN network on the 5G network. We do not modify lots of parts of the 5G architecture in reference point representation; we replace the functions of UPF with the SDN network, and the SMF can communicate with the SDN network via the existing protocol (e.g., Packet Forwarding Control Protocol (PFCP)) as shown in the round rectangle frame of Figure 6.

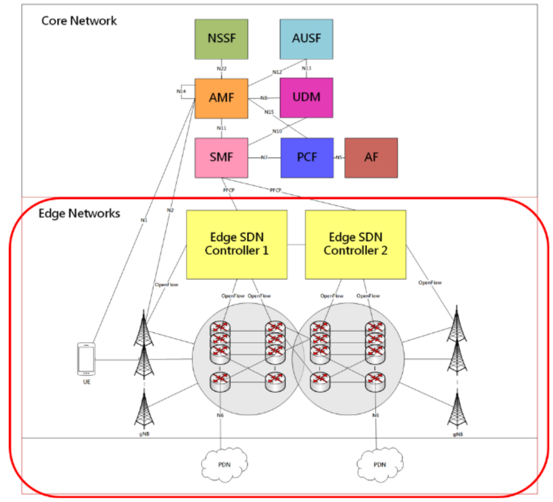


**Figure 6: ES-5G architecture**

We deploy the SDN Controller similarly to LC-PD [14] and combine it with the concept of edge computing. The SDN Controller in the edge network is called the Edge SDN Controller. All Edge SDN Controllers share information about their underlying resource via the East-West interface. In addition, they all have a global network view. When the UE sends the packet to other UE in the different edge networks, the Edge SDN Controller makes it easier to configure the forwarding path. However, there is no standard for the East-West interface.

We set up an OpenFlow Switch that connects to the external network called OpenFlow Switch (PSA); other OpenFlow Switches connect according to their port numbers. Deploying the SDN components in the edge network can shorten the forwarding distance to lower latency. We do not discuss the OpenFlow Switch deployment, such as fat-tree and leaf spins. Finally, our ES-5G network is shown in Figure 6.

We describe the functions of the ES-5G network in the following. However, we only describe the different functions in the original 5G network.

- The gNB is seen as an OpenFlow Switch, meaning the gNB has the original control plane functions of gNB and data plane functions of OpenFlow Switch. Hence, the gNB must support the OpenFlow protocol to communicate with the Edge SDN Controller. This design enables the forwarding path between gNB and OpenFlow Switch without the utilization of tunneling protocols; it avoids header encapsulation to increase the data rate. In our scenario, the gNB is not supporting the Xn application protocol to each other.
- The SMF must communicate with the Edge SDN Controller via the modified PFCP. Although the SMF is seen as the application layer for the SDN network, its primary function is forwarding the path information to the Edge SDN Controller.
- The Edge SDN Controller analyzes, manages the underlying resource, and shares its underlying resource via the East-West interface with other Edge SDN Controllers. It communicates with the SMF via the modified PFCP and the OpenFlow

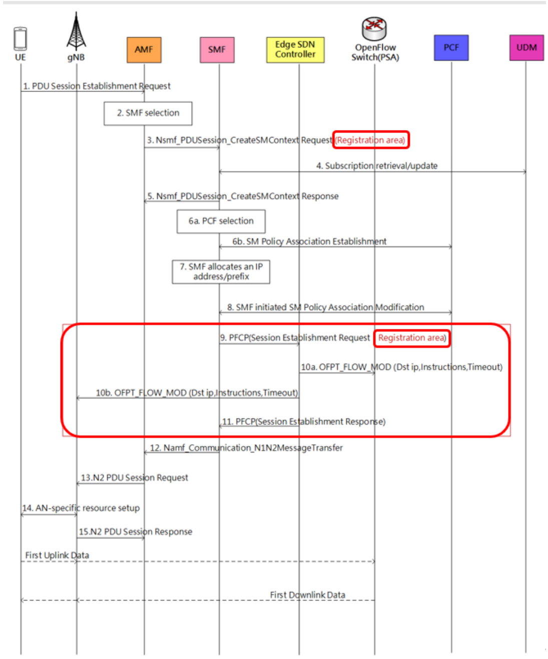


**Figure 7: PDU session establishment procedure in ES-5G**

Switches via the OpenFlow protocol. Its primary function is according to the forwarding information from the SMF to calculate the best forwarding path and modify the flow tables of OpenFlow Switches.

- The OpenFlow Switch looks up and forwards the packets according to the flow tables. It also shares its resource status with the Edge SDN Controller via the OpenFlow protocol.

## 3.2 Modified Basic Procedures

In this section, we modify the basic procedures in [18] and design the new handover procedures. The UE registration area (RA) records the common-use connection with the gNBs; the AMF produces it via the last visited Tracking Area List (TAI). The previously mentioned procedures are executed in one edge network; we do not discuss forwarding the packet between two edge networks, so it is seen as an extranet network transmission.

- PDU Session Establishment procedure

We show the PDU Session Establishment procedure in the ES-5G network as shown in Figure 7. Here, we describe and mark the different sections with the original 5G network using the round rectangle frame.

Step 3: We design the AMF to send the UE registration area to the SMF via the Nsmf_PDUSession_CreateSMContext Request message.

Step 9: Our final purpose is sending the UE registration area to the Edge SDN Controller to calculate and pre-establish the best forwarding path between the gNBs and OpenFlow Switch (PSA), so the SMF sends the UE registration area to the Edge SDN Controller and notifies the Edge SDN Controller session establishment via the modified PFCP in this step.

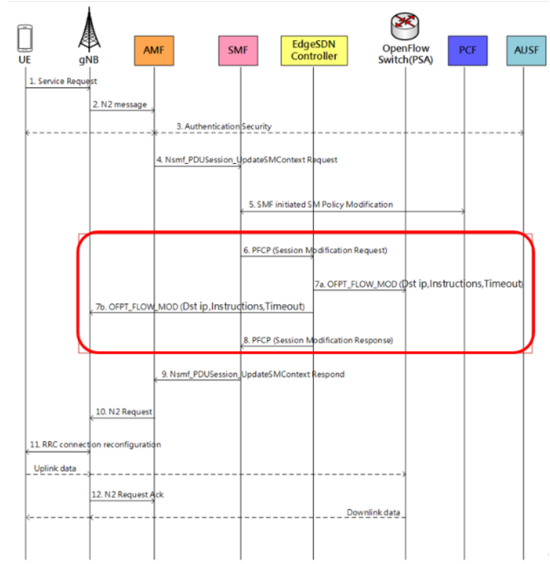


**Figure 8: UE-triggered service request procedure in ES-5G**

Step 10a: The Edge SDN Controller modifies the OpenFlow Switches' flow entries to establish the DL forwarding path.

Step 10b: The Edge SDN Controller modifies the OpenFlow Switches' flow entries to establish the UL forwarding path. The UL and the DL paths are separately established in the original network. However, in our design, the UL and DL paths are established at the same time.

Step 11: The Edge SDN Controller responds to the SM information to the SMF via the modified PFCP.

- UE-triggered Service Request procedure

We show the UE-triggered Service Request procedure in the ES-5G network, as shown in Figure 8. Here, we describe and mark the different sections with the original 5G network using the round rectangle frame.

**Step 6**: The SMF sends the message to the Edge SDN Controller to establish the session.

**Step 7a**: The Edge SDN Controller modifies the OpenFlow Switches' flow entries to establish the DL forwarding path.

**Step 7b**: The Edge SDN Controller modifies the OpenFlow Switches' flow entries to establish the UL forwarding path.

**Step 8**: The Edge SDN Controller responds to the SM information to the SMF.

- Network-triggered Service Request procedure

We show the Network-triggered Service Request procedure in the ES-5G network, as shown in Figure 9. Here, we describe and mark the different sections with the original 5G network using the round rectangle frame.

**Steps 1-2**: After the OpenFlow Switch (PSA) receives the downlink packet, it does not know how to process it, so it sends the PACKET_IN message to the Edge SDN Controller.

**Step 3**: The Edge SDN Controller forwards the packet to the SMF via the modified PFCP. The packet will be buffered in the SMF。

- Inter-RA Handover procedure

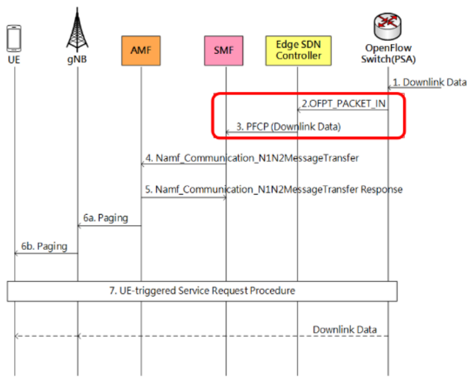


**Figure 9: Network-triggered Service Request Procedure in ES-5G**

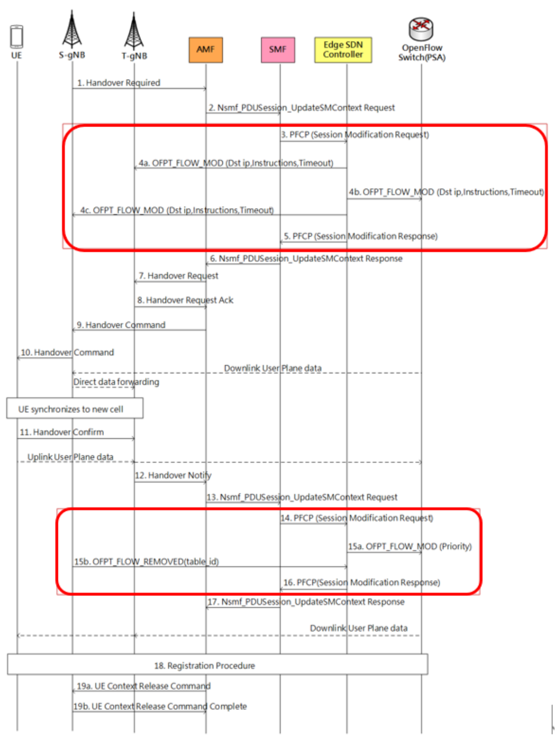


**Figure 10: Inter-RA Handover procedure in ES-5G**

We modify the N2 Handover procedure in the 5G network and rename the N2 Handover procedure to the Inter-RA Handover procedure. Finally, Figure 10 shows the Inter-RA Handover procedure. Here, we describe and mark the different sections with the original 5G network using the round rectangle frame.

**Step 3**: The SMF sends the message to the Edge SDN Controller to establish the session.

**Steps 4a-4b**: The Edge SDN Controller modifies the OpenFlow Switches' flow entries to establish the UL and DL forwarding path.

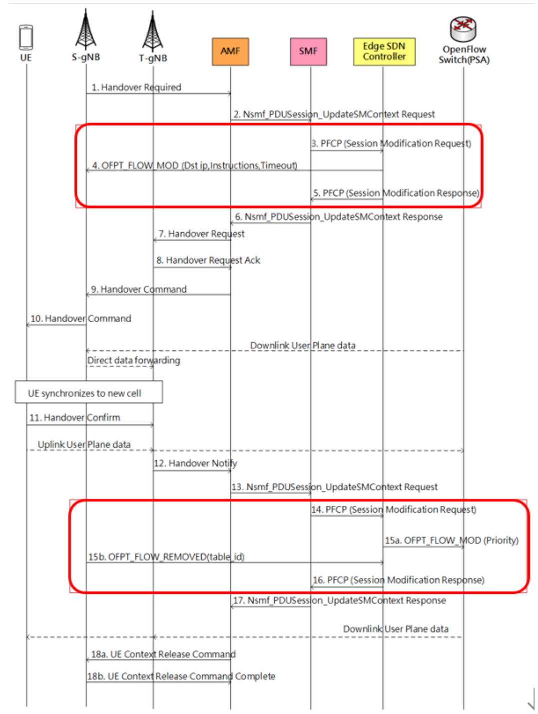


**Figure 11: Intra-RA Handover procedure in ES-5G**

**Step 4c**: The Edge SDN Controller modifies the OpenFlow Switches' flow entries to establish the indirect data forwarding path from S-gNB to T-gNB. This step can avoid packet loss during the handover procedure.

**Step 5**: The Edge SDN Controller responds to the SM information to the SMF.

**Step 14**: The same step as step 3.

**Step 15a**: The Edge SDN Controller modifies the DL session established in Step 4b to the highest priority.

**Step 15b**: When the DL session established in step 4c is "timeout," and its "flag" is configured OFPFF_SEND_FLOW_REM, the OpenFlow Switch sends the OFPT_FLOW_REMOVED message to notify the Edge SDN Controller that the session is released. In the original 5G network, setting the timer or sending the end marker can release the session.

**Step 16**: The same step as step 5.

- Intra-RA Handover procedure

Figure 11 shows the intra-RA handover procedure in the ES-5G network. Again, we refer to [18, 20, 21] to design this procedure.

The forwarding paths between the gNBs and the OpenFlow Switch (PSA) are established in the PDU Session Establishment procedure. To avoid packet loss, we must establish the indirect data forwarding path from S-gNB to T-gNB (step 5). Therefore, another round rectangle frame is the same as the Inter-RA Handover procedure.

We effectively pre-establish forwarding paths, such as control latency, to reduce the control signal overhead. This procedure is similar to the Intra-RA Handover procedure. However, the Edge

**Table 1: Function comparison UPF with SDN network**

| UPF functions [17] | SDN network functions [19, 20] |
|---|---|
| Anchor point for Intra-/Inter-RAT mobility | Anchor point for Intra-/Inter-RAT mobility |
| Allocation of UE IP address/prefix | N/A |
| External PDU Session point of interconnect to Data Network | External PDU Session point of interconnect to Data Network |
| Packet routing forwarding | Packet routing forwarding |
| Packet inspection | Packet inspection |
| QoS handling for data plane | QoS handling via the "meter table." |
| Transport level packet marking in the uplink and downlink | |
| Uplink Traffic verification | |
| Data Plane part of policy rule enforcement | Data Plane part of policy rule enforcement |
| Traffic usage reporting | OpenFlow Switches state reporting |
| Downlink packet buffering and downlink data notification triggering | N/A |
| Sending and forwarding of one or more "end markers" to the S-gNB node | Sending and forwarding of one or more "end markers" to the S-gNB node or setting up the "timeout" to release the session |
| Functionality to respond to ARP requests and/or IPv6 Neighbor Solicitation requests | Functionality to respond to ARP requests and/or IPv6 Neighbor Solicitation requests |

SDN Controller processes fewer things than the Inter-RA Handover procedure.

## 3.3 SDN Network Function Comparison

In this section, we describe why we can replace the 5G UPF with the SDN network and the packet processing flow of the SDN network. We first list the functions of the 5G UPF components and describe how the SDN network replaces them. Finally, Table 1 shows the comparison table between them.

- The UPU is a root node of packet forwarding called an anchor point for Intra-/Inter-RAT mobility in the original 5G network. In our design, the SDN network also has an anchor point called OpenFlow Switch (PSA).
- In the original 5G network, IP address allocation has three mechanisms. The first one is the UPF allocates the IP address according to the necessary information from the SMF; the second one is the SMF chooses and allocates the IP address from the IP address pool; the last one is the external data network such as DHCP/DN-AAA server assigns the IP address. The SDN network does not have this function in our design, but we still have the second and last mechanisms to allocate the IP address.
- The original 5G network has a UPF (PSA) that can connect to the external network; in our design, the OpenFlow Switch (PSA) has the same function.
- The UPF forwarding packet via the traffic forwarding methods from the SMF in the original 5G network; in our design, the SMF sends the Forwarding Action Rule (FAR) information to the Edge SDN Controller via the modified PFCP, and the Edge SDN Controller will calculate and configure the best forwarding path to the OpenFlow Switches. Then, the OpenFlow Switches forward the packets via the flow tables.
- When the packet arrives at the UPF (PSA), the UPF (PSA) inspects and filters the packet to the service data flow with the corresponding QoS policies such as priority, bandwidth control, etc. Then, the UPF(PSA) sends the packet to the UE according to the QoS policies, as shown in Figure 12.

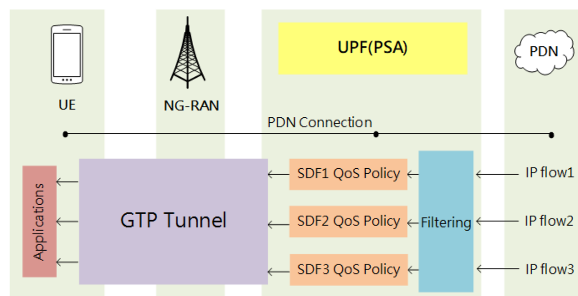


**Figure 12: QoS enforcement in 5G network**

- The SMF configures the QoS parameters (e.g., 5G QoS identifier /guaranteed flow bit rate) to the UPF; it enables the UPF to mark the packet, such as the Differentiated Service Code Point (DSCP) value in the IP header of the packet. Furthermore, if the UE sends the packet to the PDN, the UPF (PSA) verifies whether the uplink packet's QoS Flow ID (QFI) matches the QoS rules.
- Figure 13 presents QoS enforcement in our design. After the OpenFlow Switch receives the packet, it looks it up. It has the corresponding meter table to process the QoS operation. The OpenFlow Switch also can mark the DSCP value in the IP header of the packet.
- The UPF enforces the policy rules of the data plane from the SMF in the original 5G network. In our design, the SMF also sends the policy rules to the Edge SDN Controller to configure the forwarding path.
- The UPF reports traffic usage to the original 5G network SMF. In our design, if the OpenFlow Switch changes its status, it will send a message to notify the Edge SDN Controller.

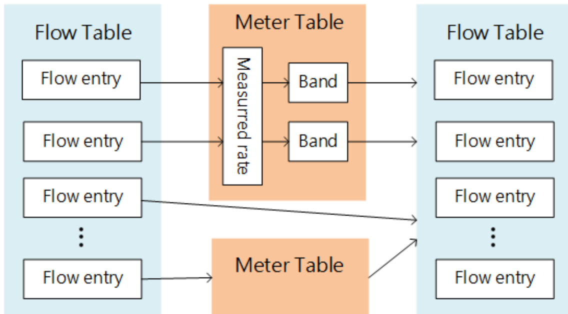


**Figure 13: QoS enforcement in SDN network**

The Edge SDN Controller knows the consistent status of its underlying resources.

- Suppose a downlink packet is sent to the UE, but the UE status is CM_IDLE. In that case, the UPF will buffer the packet and trigger the downlink data notification to SMF. However, in our design, the OpenFlow Switch (PSA) cannot buffer the packet, so we forward the packet from the Edge SDN Controller to the SMF to buffer it, as shown in Figure 9.
- The UPF sends the end marker packet to release the session of S-gNB and assist the reordering function in the T-gNB, such as the Xn-based Handover procedure [18]. In our design, after the Edge SDN Controller configures the new forwarding path and responds to SMF, the SMF constructs the end marker packet(s) and sends it to the Edge SDN Controller. Then, the Edge SDN Controller sends the end marker packet to the OpenFlow Switches to notify the S-gNB to release the session and notify the T-gNB to reorder the packets. In addition, we also can set up the "timeout" in the flow entry to release the session. When the "timeout" is exceeded, the OpenFlow Switches have a flow expiry mechanism to remove the flow entries. Furthermore, the Edge SDN Controller may actively remove flow entries by sending delete flow table modification messages (e.g., OFPFC_DELETE) to the OpenFlow Switches.
- In the original 5G network, two methods exist to respond to the Address Resolution Protocol (ARP) requests or IPv6 Neighbor Solicitation requests. The first method is the UPF (PSA) responds to the request according to its local cache information, i.e., the mapping between the UE MAC address to the UE IP address; the second method is the UPF(PSA) redirects the ARP traffic to the SMF, and the SMF will respond to the request according to its local cache information. In our design, when the Edge SDN Controller recovers the ARP request, the Edge SDN Controller resends a new ARP request to the OpenFlow Switches in the same domain and waits for the corresponding OpenFlow Switch to respond.

## 3.4 SDN Packet Processing Flow

In this section, we describe the packet processing flow of the OpenFlow Switch [20] in the SDN network, as shown in Figure 14.

The method of processing packets in the OpenFlow Switch is called pipeline processing. When the packet arrives at the OpenFlow

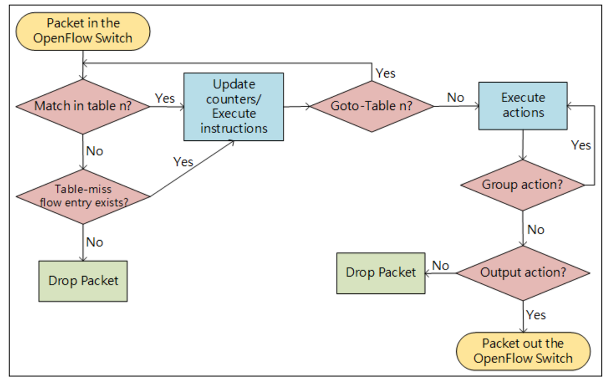


**Figure 14: Packet processing flowchart in OpenFlow Switch**

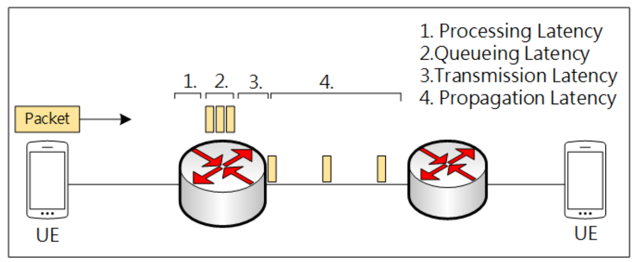


**Figure 15: End-to-end latency**

Switch, the OpenFlow Switch matches the packet with the "match field" in the flow table. For example, the matching methods may match the packet header fields, the packet ingress port, etc.

The packet executes the "instructions" if a matching entry is found. The OpenFlow Switch executes the "table-miss" entry if no match is found. The "table-miss" entry describes how to process the no-match packet; for example, the packet may be dropped, forwarded to the Edge SDN Controller via the PACKET_IN message, or continue to the following table. The packet will be dropped if there is no "table-miss" entry in the flow table.

If the flow entry is with the Goto-Table instruction, the packet will be processed according to the previous method; if there is no Goto-Table instruction, the OpenFlow Switch will execute the "actions." Then, the OpenFlow Switch verifies whether the Group action/output action was executed. Finally, the output action describes the output port in the OpenFlow Switch.

The flow table may have the ingress and egress tables that process the packet ingress and egress. However, Figure 14 assumes no egress table to process the packet.

## 4 THEORETICAL ANALYSIS

This section analyzes the end-to-end latency forwarding packets from the source UE to the destination UE. End-to-end latency includes the processing, transmission, propagation, and queueing latency, as shown in Figure 15.

**Table 2: The number of 5G and ES-5G components**

| Component | Notation | Number (#) |
|---|---|---|
| UE | $N_{UE}$ | 10 |
| gNB | $N_{gNB}$ | 10 |
| Router | $N_{router\ backhual}$ | 16 |
| UPF | $N_{UPF}$ | 5 |
| UPF(PSA) | $N_{UPF\ (PSA)}$ | 1 |
| OpenFlow Switch | $N_{OFS\ backhual}$ | 16 |
| OpenFlow Switch (PSA) | $N_{OFS\ (PSA)}$ | 1 |

## 4.1 End-to-end Latency

We assume the forwarding devices are connected with fiber lines, and the distance between the forwarding devices is the same in the 5G network and ES-5G network. The speed of light has negligible latency because its data rate is so fast (3*108 m/s). Therefore, the propagation latency does not calculate it.

The transmission latency means the forwarding latency from the forwarding devices' ports to the fiber lines. The bandwidth of the fiber lines is the main factor to affect the transmission latency. Here, we assume the bandwidth is the same, so we do not calculate the transmission latency.

The packet forwarding is via the UPF and routers in the 5G network; in our design, the packet forwarding is via the OpenFlow Switches.

Table 2 shows our environmental parameters. We set up the one gNB that serves the 10 UEs [14], and the number of the gNBs is 10. An equal number between the gNB to the UPF or the gNB to the OpenFlow Switch (PSA) is fairer to compare them. The routers and OpenFlow Switches are in the backhaul network, and they are difficult to define the number of them. Therefore, we assume the average number of forwarding packets via the gNB to the UPF or the gNB to the OpenFlow Switches (PSA) is 4. However, we set up the number of UPFs as five because we assume that the range of the core network is composed of five edge networks.

## 4.2 Operating Result

This section describes the end-to-end latency of the 5G and ES-5G architectures.

We calculate the total end-to-end latency time in 5G and ES-5G architectures using MATLAB software. Figure 16 shows that the ES-5G architecture has lower latency than the original 5G. We describe the reason for the result; the first reason is that the OpenFlow Switches forward the packets according to the flow tables. However, the UPFs and the routers need to decide how to forward the packets; they need more time to process them. The second reason is that our SDN network is configured at the edge network; this configuration can shorten the forwarding distance to reduce the latency.

The result shows the steady-state situation. When the arrival rate approaches the service rate, the ratio will soar.

## 4.3 Control Plane Message Exchanges

We analyze the message exchanges of the handover procedures, the N2 handover procedure in the original 5G network, the Inter-RA Handover procedure, and the Intra-RA Handover procedure in the ES-5G network.

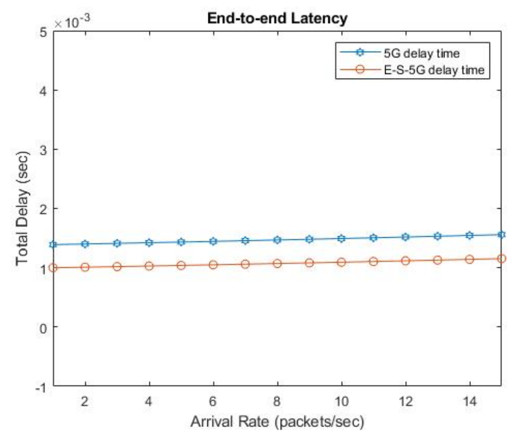


**Figure 16: Comparison of total latency time in 5G and ES-5G architectures**

Table 3 shows the operation result of the handover procedures. This result is that the handover procedures in the ES-5G network have fewer message exchanges than in the original 5G network; it proves our design is valid to reduce the handover procedure's signaling cost and processing time.

However, the Inter-RA Handover procedure is similar to the Intra-RA Handover one. We use one message to present the SDN Controller modifies the underlying OpenFlow Switches. The number of OpenFlow Switches affects the total handover processing time of the SDN Controller. For example, the flow tables need to be modified if we have five OpenFlow Switches (four OpenFlow Switches and one gNB are passed by the packet) [22]. The SDN Controller's processing time is 3 $\mu s$, so the total variation processing time, including DL and UL paths, is 3*5*2 $\mu s$. This variation proves that pre-establishing the forwarding path can reduce the handover processing time.

# 5 SIMULATION

In this section, we simulate the network environment via the OMNeT++ software. The result can verify that our analysis of end-to-end latency in Section 4 is correct.

The OMNeT++ software [23] is a simulation platform programmed in C++. Our simulation project includes three main files:

**Table 3: The number of message exchanges of handover procedures**

| | Number of message exchanges |
|---|---|
| N2 Handover | 22 |
| Inter-RA Handover | 19 |
| Intra-RA Handover | 18 |

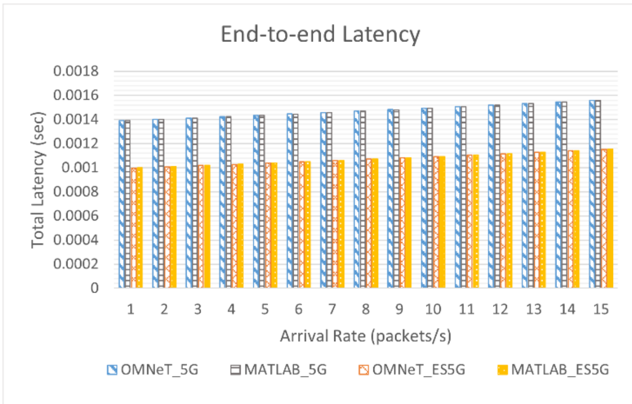


**Figure 17: Comparison of total latency time in OMNeT++ and MATLAB software**

a NED file, a C++ file, and an INI file. The NED file defines the network topology. We can assemble into more significant components and models using a high-level language called NEtwork Description (NED) language or the Graphical User Interface (GUI). The C++ file describes the behavior of the components. Finally, the INI file is used to set up the simulation parameters.

We generate packets via the Poisson distribution with an arrival rate ($\lambda$) to simulate user packet forwarding. A UE generates the ten times packets and sends them to the gNB because one gNB serves the ten UEs. After the gNB receives the packets, it sends the two-point-five times packets to the backhaul network. After the UPF gets the packets, it sends the five times packets to UPF (PSA) to satisfy our environment.

Figure 17 shows the simulation result in OMNeT++. We compare this result with our analysis of end-to-end latency formulas. We can see this result approached the same as our analysis, so it is efficient to verify that our end-to-end latency is correct.

## 6 CONCLUSION

In this paper, we replace the data plane of the existing 5G architecture with the SDN technology and combine the concept of edge computing to propose a new architecture called the ES-5G network.

We modify the basic procedures in the 5G network and propose a new mechanism to reduce the processing time in the handover procedure. First, we offer the new handover procedures, the Intra-RA Handover and Inter-RA Handover procedures. After that, we describe the function comparison between the UPF component that is responsible for packet forwarding and the SDN network in ES-5G architecture to verify that our replacement method can maintain the 5G regular operation. By the way, we also describe the packet processing flow of the OpenFlow Switch in ES-5G.

Then, we calculate the end-to-end latency between the ES-5G network and the original 5G network and the message exchanges of the handover procedures to prove that our design has a lower control overhead and more downprocessing time. Finally, we simulate the previous evaluation to verify that our analysis is correct using the OMNeT++ software.

In future work, we plan to implement ES-5G architecture and design the shortest path algorithm in the Edge SDN Controller or combine the network slice with satisfying the different service scenarios on the 5G network. We also plan to discuss the deployment of the forwarding devices to bring lower latency.

## ACKNOWLEDGMENTS

This work was partly sponsored by the Ministry of Science and Technology of Taiwan under grant MOST111-2221-E-194-019-.

## REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," IEEE Commun. Surveys Tuts., vol. 18,no. 3, pp. 1617–1655, 3rd Quart., 2016.
[2] A. Huang, N. Nikaein, T. Stenbock, A. Ksentini, and C. Bonnet, "Low latency MEC framework for SDN-based LTE/LTE-A networks," in Proc. of the IEEE International Conference on Communications, 2017.
[3] S. Kekki *et al.*, "MEC in 5G networks," Sophia Antipolis, France, ETSI, White Paper, Jun.2018.Accessed: Dec.4,2018.[Online].Available:https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf
[4] M. Moradi, Y. Lin, Z. M. Mao, S. Sen, and O. Spatscheck, "SoftBox: A customizable, low-latency, and scalable 5G core network architecture," IEEE J. Sel. Areas Commun., vol. 36, no. 3, pp. 438–456, Mar. 2018.
[5] R. Trivisonno, R. Guerzoni, I. Vaishnavi, and D. Soldani, "Towards zero latency software defined 5G networks," in Proc. IEEE Conf. Commun. Workshop (ICCW), London, U.K., pp. 2566–2571, Jun. 2015.
[6] A. Jain, N. Sadagopan, S. K. Lohani, and M. Vutukuru, "A comparison of SDN and NFV for re-designing the LTE packet core," in Proc.IEEE Conf. Netw. Function Virtual. Softw. Defined Netw. (NFV-SDN), Palo Alto, CA, USA, pp. 74–80, 2016.
[7] M. T. Raza, S. Lu, M. Gerla, and X. Li, "Refactoring network functions modules to reduce latencies and improve fault tolerance in NFV," IEEE J. Sel. Areas Commun., vol. 36, no. 10, pp. 2275–2287, Oct. 2018.
[8] T. Ochiai, K. Matsueda, T. Kondo, H. Takano, R. Kimura, R. Sawai, and F. Teraoka, "MocLis: a non-tunneling moving cell support protocol based on Locator/ID split for 5G system," IEEE International Conference on Communications (ICC), pp. 1–7, 2018.
[9] H. Ko, I. Jang, J. Lee, S. Pack, and G. Lee, "SDN-based distributed mobility management for 5G," in Consumer Electronics (ICCE), 2017 IEEE International Conference on. IEEE, pp. 116–117, 2017.
[10] H. Jin, Y. Jin, H. Lu, C. Zhao, M. Peng, "NFV and SFC: a case study of optimization for virtual mobility management," IEEE J. Sel Areas Commun., vol. 36, no. 10, pp. 2318-2332, Oct. 2018.
[11] S. Kuklinski, Y. Li, and K. T. Dinh, "Handover management in SDN-based mobile networks," 2014 IEEE GLOBECOM Wksps., Austin, TX, pp. 194–200, 2014.
[12] M. Erel-Özçevik, B. Canberk," Road to 5G reduced-latency: A software-defined handover model for eMBB services," IEEE Trans. Veh. Technol., vol. 68, no. 8, Aug. 2019.
[13] A. Gasmelseed, R. Ramalakshmi, "Performance analysis of centralized and distributed SDN Controllers for load balancing application," in 2nd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, 2018.

[14] C. N. Tadros, M. R. M. Rizk, and B. M. Mokhtar," Software-defined network-based management for enhanced 5G network services," IEEE Access, vol. 8, pp. 53997-54008, Mar. 2020.
[15] F. Bannour, S. Souihi, and A. Mellouk, "Distributed SDN control: Survey, taxonomy, and challenges," IEEE Commun. Surveys Tuts., vol. 20, no. 1, pp. 333–354, 1st Quart., 2018.
[16] M. Jammal, T. Singh, A. Shami, R. Asal, and Y. Li, "Software-defined networking: State of the art and research challenges," Computer Network, vol. 72, pp. 74–98, Oct. 2014.
[17] 3GPP TS 23.501: "System Architecture for the 5G System (5GS); Stage 2," V16.2.0, Release 16, Sep. 2019.
[18] 3GPP TS 23.502: "Procedures for the 5G System (5GS); Stage 2," V16.2.0, Release 16, Sep. 2019.
[19] ONF TR-521: "SDN Architecture," Issue 1.1, 2016.
[20] ONF TS-025: "OpenFlow Switch Specification," Version 1.5.1, Mar. 26, 2015.
[21] 3GPP TR 38.331: "NR; Radio Resource Control (RRC); Protocol specification," V15.7.0 Sep. 2019.
[22] J. Prados-Garzon, O. Adamuz-Hinojosa, P. Ameigeiras, J. J. RamosMunoz, P. Andres-Maldonado and J. Lopez.Soler, "Handover implementation in a 5G SDN-based mobile network architecture," 27th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2012), Valencia, Spain, 3-5, Jul. 2012.
[23] OMNeT++ Tutorials. https://docs.omnetpp.org/