Smart Virtualization for Packet Forwarding in 5G and Beyond Communication Networks

By

FOUAD A. YASEEN

Supervisor: **Professor Hamed Al-Raweshidy**

Department of Electronic and Computer Engineering
College of Engineering, Design and Physical Sciences
Brunel University London
United Kingdom

This thesis is submitted for the degree of
Doctor of Philosophy (PhD) in Communications

October 2019

Abstract

In this thesis, novel ideas have been proposed to tackle the delay and connection continuity, which are caused by different factors related to wired and wireless communication system networks. The vast majority, if we do not say all of these systems adopt packet switch schemes to transfer the data amongst network devices. Moreover, all these systems aim to deliver the data from the source to the destination according to particular Identifiers (IDs) of these devices. The most well-known IDs that are used to distinguish devices are IP address, MAC address, and Subscriber National Number.

By virtualizing the servers of communication systems, new concepts of communication networks have emerged. For instance, a mobile operator's core network function servers could be virtualized, installed, and run by Virtual Machines (VMs) to execute these functions. Also, routers, switches, firewalls, and other network devices could be virtualized by using SDN, NFV, and new- generation protocols such as OpenFlow to create a high performance of the virtualized communication network. From this point of view, we proposed novel concepts such as SVeNB which mimicked the functions of the base station and the core network of a mobile operator network. The SVeNB performance exceeded the C-RAN with 68% in user profiles treatment. Also, the SVeNB reduced the End-to-End delay to 62%. Other advantages of the virtualization are ease of separating the functions and control layer of the networks. This approach urged researchers to suggest a new communication network topologies and innovative designing of flexible and programmable routing protocols.

As a result of these approaches, emerging the SDN networks to carry packet forwarding schemes on the communication networks. Based on the facts mentioned above, we designed a novel idea to generate a tag as a mobile node's ID from E.164 standard numbering and MAC address to handle the packets inside the networks. The results showed that the packet loss rate decreased to 4% of that were lost during the handover delay time or while packets re-direction mechanism. At the same time, the MN could receive 96.4% of the data that was lost during the handover process.

Mobility management is a vital issue in wireless communication, due to the necessity of changing the ID of the wireless attached Access Points (APs) by a moving target which connects to that APs. The biggest obstacle of mobility is that the addresses resolution should be made in real time. More difficulty is added when the motion of moving targets is very speedy, for example, such as High-Speed Trains. A novel proactive scheme has been presented by chapter 4 for directing the packet flows among the APs, with support of the trigger signal to activate layer 2 handover. By using the triggering signal, the performance of the suggested network surpassed the performance results that were not supported by the triggering signal. The average control delay time was reduced by nearly 45% and the retrieved data were roughly 90% of packet loss when adopting the triggering signal system.

To the cause of my existence in this life, my parents.

To all my sisters and brothers.

To my lovely children.

To all my friends.

To my wings,
my wife, with her, I will have been flying high in the sky.

Declaration

I declare that this thesis is my own work and is submitted for the first time to the Post-Graduate Research Office. The study was originated, composed and reviewed by myself and my supervisors in the Department of Electronic and Computer Engineering, College of Engineering, Design and Physical Sciences, Brunel University London, UK. All the information derived from other works has been properly referenced and acknowledged.

By
FOUAD A. YASEEN
October 2019

Acknowledgements

First of all, my gratefulness and appreciation would go to the Ministry of Higher Education and Scientific Research (MOHESR) / Iraq, Iraqi Cultural Attache / London, and University of Baghdad / Iraq for funding and supporting this study.

I also would like to give my sincere thankfulness to Professor Hamed Al-Raweshidy for his valuable advice, encouragement, and guidance. His endless support, and helpful reviews. Most appreciated thing is his perception and patience to put this study in the right way.

Last but not least, I appreciate my family role for their support, love, encouragement and inspiration to me through the times of my study.

Table of contents

List of figures

List of tables

List of Abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AaaS	Anything as a Service
AOFS	Aggregator OpenFlow Switch
AP	Access Point
APc	Current Access Point
API	Application Program Interface
APn	Next Access Point
AR	Access Router
AS	Access Stratum
B.W	Bandwidth
BBU	Baseband Unit
BSC	Base Station Controller
CAM	Content Addressable Memory
CAN	Content Addressable Network
CAPEX	Capital Expenditure
CC	Cloud Computing
CN	Core Network
CoA	Care of Address

C-RAN	Cloud Radio Access Network
DHT	Distributed Hash Table
DP	Data Plane
DTX	Discontinuous Transmission
EID	Endpoint Identifier
EIGRP	Enhanced Interior Gateway Routing Protocol
eNB	evolved Base Station
EOFS	Edge OpenFlow Switch
EPC	Evolved Packet Core
EPS	Evolved Packet System
ESDNc	Edge SDN Controller
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
F/TDMA	Frequency / Time Division Multiple Access
FA	Foreign Agent
FER	Frame Error Rate
FPGA	Field-Programmable Gate Array
GPS	Global Positioning System
GSM	Global System for Mobile
HA	Home Agent
HARQ	Hybrid Automatic Repeat Request
HID	Host Identifier
HIP	Host Identifier Protocol
HoA	Home Address
HSR	High-Speed Railway
HSS	Home Subscriber Server
HST	High-Speed Train
IaaS	Infrastructure as a Service
ICI	Intercarrier Interference
ID	Identifier

IID	Interface Identifier
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
IS-IS	Intermediate System - Intermediate System
ITU-T	International Telecommunications Union for Telecommunication
LAN	Local Area Network
LISP-DMC	Locator-Identifier Separation Protocol Distributed Mobility Control
LISP-host	Locator-Identifier Separation Protocol Host
LMA	Local Mobility Anchor
LMS	Local Map Server
LSDNc	Local SDNc
LTE	Long-Term Evolution
MAC	Media Access Control
MAN	Metropolitan Area Network
ME	Mobile Equipment
MIMO	Massive Multiple Input multiple Output
MIP	Mobile IP
MME	Mobility Management Entity
mmW	Millimeter Wave
MN	Mobile Node
MOFI	Mobile-Oriented Future Internet
MT	Mobile Terminal
MTU	Maximum Transmission Unit
NAT	Network Address Translation
NFS	Network Functions Slicing
NFV	Network Functions Virtualization
NIC	Network Interface Card
NID	Network Identifier

NIST	National Institute of Standards and Technology
NS	Network Slicing
OFDM	Orthogonal Frequency Division Multiple
OFDMA	Orthogonal Frequency Division Multiple Access
OFS	OpenFlow Switch
ONF	Open Networking Foundation
OPEX	Operational Expenditure
OS	Operating System
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First
OUI	Organizationally Unique Identifier
PaaS	Platform as a Service
PCEF	Policy Control Enforcement Function
PCRF	Policy and Charging Rules Function
PDN	Packet Data Network
P-GW	Packet Gateway
PRACH	Physical Random Access Channel
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RACH	Random Access Channel
RAP	Random Access Preambles
RAR	Random Access Response
RA-RNTI	Random Access Radio Network Temporary Identity
RF	Radio Frequency
RRH	Remote Radio Head
RIP	Routing Information Protocol
RLC	Radio Link Control
RLOC	Routing Locator
RNC	Radio Network Controller

RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RTT	Round Trip Time
SaaS	Software as a Service
SAE	System Architecture Evolution
SDN	Software-Defined Networking
SDR	Software-Defined Radio
S-GW	Serving Gateway
Shim6	Site Multihoming by IPv6 Intermediation
SID	Subnetwork Identifier
SN	Sender Node
SNN	Subscriber National Number
SVeNB	Smart Virtual eNB
TB	Transport Block
TE	terminal Equipment
TR	Tunnel Router
TSS	Transmitted Signal Strength
TTI	Transmission Time Interval
UMTS	Universal Mobile Telecommunication System
USIM	UMTS Subscriber Identity Module
VLSM	Variable Length Subnet Mask
VM	Virtual Machine
WAN	Wide Area Network
WiFi	Wireless Fidelity
WiGig	Wireless Gigabit
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless LAN

Chapter 1

Introduction

1.1 State-of-the-Art

5G communications network technologies should be characterized by the high throughput rate ranging (10 to 100 Gbps) for backhaul networks, delivered data rate (100 Mbps to 1Gbps) for an end user, the End-to-End control latency less than 1 ms, quicker configuration time, reduced control signaling overhead, and less energy consumption [1]. These features should have existed for all 5G networks whether it was a wired, wireless, mobile, or stationary networks. The 5G networks' infrastructure should also include flexible techniques to handle the data packets that pass through it. Most promising enabling technologies that are demanded to be selected by 5G networks structure are cloudification, virtualizing, software-defined, and slicing. These different concepts have a wide diversity of innovating adaptable ideas to design, virtualize, and manage the 5G networks [2]. Based on the support of these technologies, the aforementioned 5G network requirements can be realized. This realization of 5G can be achieved via using Software-Defined Networking (SDN), Network Functions Virtualization (NFV), Network Slicing (NS), Network Functions Slicing (NFS), and cloudification [3].

SDN networks depend on the separation of the network Control Plane (CP) from the user Data Plane (DP), providing direct programming capability for network control functions. SDN is a complementary technology for NFV because it can provide the foundation on which virtual network functionality can run. The SDN design relies on logically centralized controllers that are software-based to gain a global network view and allowing the network to seem logically as a single virtual switch for policy rules and applications.

The NFV technology allows networks' operator to implement network functions in software running on general-purpose computing platforms to meet the 5G demands. The NFV executes network functions by software instead of dedicated hardware. This hardware migration to software running in a cloud environment, the NFV leads reduction in both Capital Expenditure(CAPEX) of the equipment cost and Operational Expenditure (OPEX) cost. Moreover, services can also be deployed, scaled up/down, more flexible, and lower delay.

Cloudification has been used to create the cloud computing centers which can be real or virtual infrastructures to perform computing processes. The principal emergence of cloudification in 5G mobile networks is Cloud Radio Access Network (C-RAN). The main concept of the C-RAN is to gather BaseBand Unit (BBU) from many Base Stations (BS)s into BBU Pool at centralized processing while leaving the antennas and the Remote Radio Heads (RRH)s at the BS places.

Following the above discussions, the integrated 5G architectures based on SDN, Network Virtualizing, and NFV technologies are flexible and scalable to enable a diversity of 5G networks such as mobile and stationary communication networks.

1.2 Motivations

As a complement to the empirical successes of previous studies in the area of communications network technologies that discussed later in chapter two, this thesis provides the principal of the motivations that have inspired this research. The motivations of the first contribution are:

1. Reducing end to end delay, according to scientific studies, the largest time delay occurs because of exchanging of control plane messages between UE and the attached network at initializing the connection session.

2. Employing network virtualizing, by sharing, slicing, and virtualizing CN functions and bringing these virtualized, shared and sliced functions close to the UEs

The second contribution motivations are:

1. Using a novel identity for a mobile node to handle the data packets during the mobility of the mobile node based on SDN, NFV, and virtualization.

2. Maintaining continuity of a session connection without interrupting the association as an MN moves between network access points.

The third contribution motivations are related to high-speed trains issues such as:

1. Obtaining the advantages of the proactive scheme by the reduction of controlling delay messages between the SDN controller and underneath OpenFlow switches.

2. Decreasing the number of lost packets during the handover procedure for both cases of increasing in the channel bandwidth capacity and the train speed.

These the main motivations that urge us to create and find creative and innovative ideas to reach our goals.

1.3 Aim and Objectives

This section presents the aim and the objectives of this thesis. The author considers different types of delay challenges associated with the 5G communication networks by addressing the delay of control layer functionalities.

1.3.1 Aim

The principal goals of this thesis are that, decrease the End-to-End CP delay, reduce the number of packets lost, and keep a session continuity. These obstacles have resulted due to processing exchanged control messages between different types of servers. Realizing these goals can be achieved by exploiting the skills of SDN, NFV, and network virtualization in the diverse communication networks to design entire programmable network architectures. Furthermore, to improve and support the scalable, controllable, and flexible management of 5G communication networks and beyond [3].

1.3.2 Objectives

The purpose of this research is to developing current communication networks and providing new and novel ideas to support different 5G communication networks. Detailed objectives are as following:

- To reduce the End-to-End control delay through decreasing the number of CP messages. For a practical example, you have to wait more than 3000 ms to hear ringing the destination phone before you say *"HELLO"*?!.
- To reduce packets loss during mobility and high-speed mobility of objects. Most of us notice when we traverse between places, and we are in the telephone conversation there is an interruption in the conversation.
- To mitigate the burden on the Core Network(CN), by partitioning, slicing, sharing, and virtualizing the main functions that involve in making decisions of an End-to-End connection.
- To reduce power consumption, through using several virtual servers within a single physical server.
- To design networks have scalable, programmable, and flexible architectures and management.

1.4 Contributions

Research contributions have provided for improving the performance of 5G communications networks concerning the End-to-End delays and which problems cause these delays. The contributions of this thesis are summarized and presented by the following:

- Proposing an innovative virtual BS that has the ability to handle and maintain a connection session without contacting the BBU and the CN, the development of the suggested BS is based on an implementation of SDN, NFV, and virtualization where a logical and smart virtualization are used to emulate the kernel of controlling functions.
- Implementing a novel identifier for a mobile node to be used as the node identifier, the SDN controller uses this identifier in making decisions govern the underneath OpenFlow switches to forward the data.

- Implementing a novel scheme for forwarding and handling the data packets during a horizontal handover of mobile nodes (focusing on data packets steering and directing within the network).
- Emulating and implementing a novel proactive scheme to forward and handle the high data rate in high-speed trains communication networks, the network has been emulated by using Mininet, SDN, NFV, Python, and MATLAB platforms.

1.5 Thesis Structure

This thesis comprises six chapters besides the references. It starts with an introductory chapter and ends with a conclusions and future works chapter. The brief description for every chapter is given by following:

- **Chapter 1** introduces the State-of-the-Art, Motivation, Aim and Objectives, Contributions, Thesis Structure, and Publications.
- **Chapter 2** presents a brief Background and Literature Review explanation of communication networks and its devices. It also illustrates some of the terms and technologies applied in the communication networks. Moreover, it reviews the evolution of mobile networks and how they merged with communication networks.
- **Chapter 3** shows the contribution that proposing a novel locally semi-standalone Smart Virtual evolved Base Station (SVeNB) for 5G mobile communication. Furthermore, it displays which of the EPS element functions are virtualized by the SVeNB.
- **Chapter 4** illustrates an innovative approach to create a tag as mobile node identity. This tag is generated by combining the E.164 standard numbering and MAC address of the mobile node. Also, this chapter introduces a new scheme for packet flow forwarding based on this tag, and the smart virtualization.
- **Chapter 5** provides a proposal for proactive forwarding of the high data rate networks that serve speedy vehicles such as high-speed trains. The delay, the packet flow forwarding, and

the packets loss problems as the high-speed train traverses amongst access points have been investigated by this chapter.

- **Chapter 6** is the final chapter that discusses the conclusions and the future works in this thesis. Furthermore, this chapter opens the field to expand for the next research to develop and enhance the current study.

1.6 List of Publications

1.6.1 Conferences

1. Yaseen, Fouad A., Nahlah A. Al-Khalidi, and Hamed S. Al-Raweshidy. "Smart Virtual eNB (SVeNB) for 5G mobile communication." In Fog and Mobile Edge Computing (FMEC), 2017 Second International Conference on, pp. 88-93. IEEE, 2017.

2. Yaseen, Fouad A., Raad S. Alhumaima, Wesam Al-Zubaedi, and Hamed S. Al-Raweshidy. "Modelling the power cost and trade-off of live migration the virtual machines in cloud-radio access networks." In Computer Science and Electronic Engineering (CEEC), 2017, pp. 122-127. IEEE, 2017.

3. Yaseen, F.A. and Al-Raweshidy, H.S., 2019, October. A Novel Mobile Node's Identifier for Beyond 5G SDN-based Networks. The 6th IEEE International Conference on Internet of Things: Systems, Management and Security (IOTSMS 2019). (Accepted)

1.6.2 Journals

1. Yaseen, F.A. and Al-Raweshidy, H.S., 2019. Smart Virtualization Packets Forwarding During Handover for Beyond 5G Networks. IEEE Access, 7, pp.65766-65780.

2. Yaseen, F.A. and Al-Raweshidy, H.S., 2019. Mini-Base Station: A Novel Smart Virtual eNB for 5G and Beyond Mobile Networks. IEEE Access, 7, pp.78560-78570.

3. Yaseen, F. and Al-Raweshidy, H., 2019. Proactive Forwarding of High Data Rate in Smart Virtualization Networks for High-Speed Trains. IEEE SYSTEMS, 2019.

Chapter 2

Background and Literature Review

2.1 Communication Networks

The term *Communication Networks* in this thesis generally refers to networks, which could be Local Area Network (LAN), Metropolitan Area Network (MAN), Wide Area Network (WAN), wired, wireless, mobile, and computer networks. Hence, these networks use *Packet* as a container to transfer the information and data between the source and destination nodes. Packets carry information of the source and the destination such as MAC and IP addresses which are used by network devices to handle data packets from the source to the destination. This information is used by network devices to deliver the data to the target node [4]. Directing a packet in a network is done by comparing the MAC or IP address with switching or routing tables respectively into the network devices.

Network devices forward or handle packets according to their forwarding or routing tables. This procedures are not a merely simple comparison between a MAC or IP address with the tables. In fact, it is very complex and needs a deep investigation on switching and routing mechanisms. To overcome the difficulties of understanding, planning, creating, and managing a communication network should be divided into multi-layers. The purpose of this division is that each layer provides a service to the upper layer by using services provided by the lower layer. The communication between the layers is peer to peer, i.e., every layer interacts with the equivalent layer in another device via a specific protocol. Any two adjacent layers communicate through interface links between them. The Open

Systems Interconnection (OSI) model standard describes the seven network layers, as shown in Fig. 2.1 [5]. The OSI model layers are:

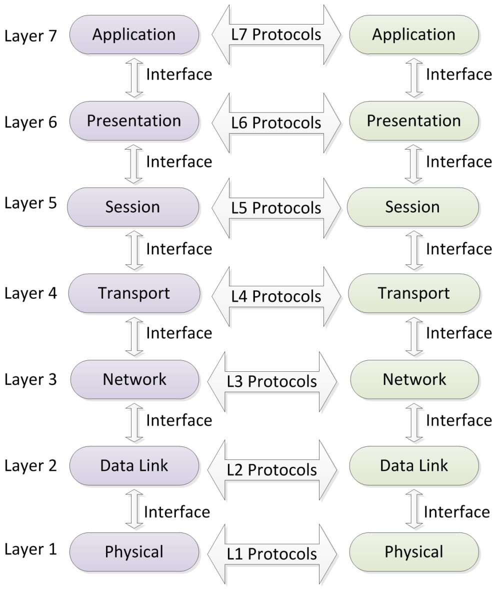


Fig. 2.1 Layers of OSI model

1. Physical Layer, it is responsible for transmitting and receiving information bits over a link, processes windows size of the connection, changes information bits to electrical signals, and synchronization at bit-level.

2. Data Link layer, it can be considered the first checkpoint of data information, adds its physical Identifier (ID) as an overhead (MAC address) to the packet to form what is known *Frame*, puts the frames on the link, detects data error and corruption through checksumming, organizes the use of the shared link, joins between the physical and the logical layers.

3. Network Layer, it is the first logical layer utilizes the logical ID (IP address) to communicate between nodes in the network, adds IP address to generate the *Packets*, deals with nodes in one network or different networks, calculates the route budget, finds the whole path between nodes, controls congestion, and forwards packets in the best path.

4. Transport Layer, it is responsible for establishing a reliable session stream between the sender and the receiver, retrieves lost packets, drops duplicated packets, re-orders received packets, and fragments and reassembles large-size messages.

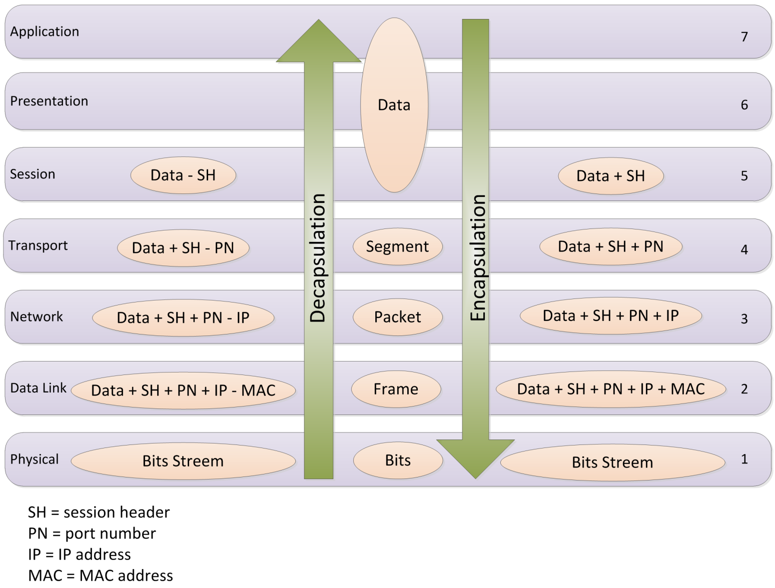


Fig. 2.2 Encapsulation and decapsulation the data

5. Session Layer, it is responsible for numbering the data segments, it provides assistance to the full-duplex reliable connection stream that is provided by the transport layer, enforces the communicated devices to follow an appropriate pattern of communication, and combines packets in groups (chains all packages are delivered in the group or none of them).

6. Presentation Layer, it agrees on data representations to deliver and format to the application layer, converse in data representation within the end-user systems, and it can be considered as data translator for the upper layer.

7. Application Layer, it provides several services such as file transfer, web surfing, network data sharing, virtual terminals, and different data and file operations.

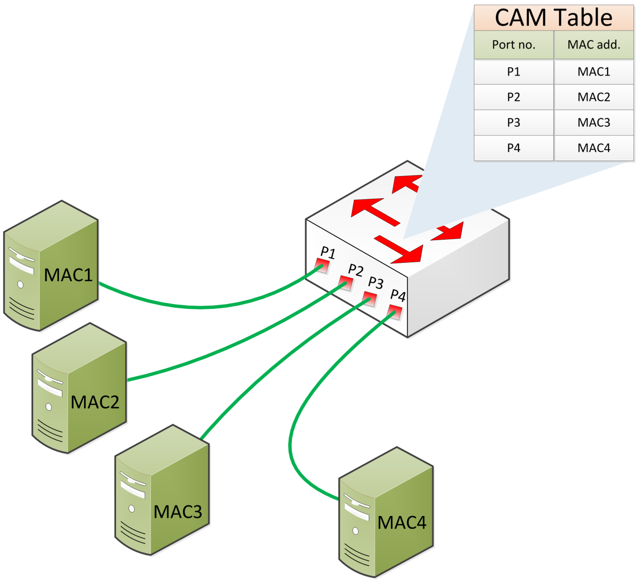
Passing the data through network layers is subjected to add or remove headers of that specific layer. The process of adding a layer header is called *Encapulation*. In contrast, removing the header is known as *Decapsulation*. Fig. 2.2 shows the changes that happen on the data.

2.2 Traditional Network Devices Switch and Router

To understand how a packet moves between network layers and into the traditional network devices, the working mechanism of these devices needs to be clarified. A brief explanation of network device's work is presented as follows:

2.2.1 Switch

Switch is a network device that can connect several nodes in a single network (LAN). It depends on destination MAC address to forward data packets between nodes. The destination MAC address is compared with the Content Addressable Memory (CAM) table which is also known as the MAC address table to direct packet on a LAN. The CAM is created by the switch so that each MAC address of a connected port number is fastened with the MAC address of the node and is listed in the CAM of the switch. Switches deal with layer 2 (data link layer), i.e., it can process MAC address only. While switches do not care about IP addresses of its connected nodes [6]. Fig 2.3 illustrate the concept of switching idea.



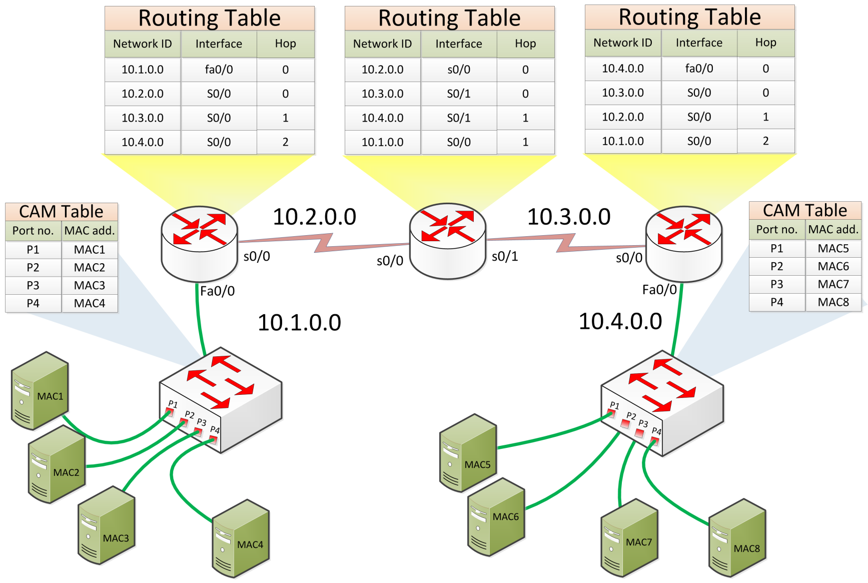
CAM Table

Port no.	MAC add.
P1	MAC1
P2	MAC2
P3	MAC3
P4	MAC4

Fig. 2.3 LAN topology based on switch device

2.2.2 Router

Router is a network device that can connect several networks. The router usually builds routing tables which are more complex and contain more information than the tables of switches, that means the router can process tasks and functions more complex and difficult than the switch can do. Not only that, but it can do all the functions provided by the switch as well as its ability to handle IP and MAC addresses. That is, a router can see the whole LAN network as one network due to its ability to deal with IP as a network identifier, furthermore its ability to handle MAC addresses. In other words, the router can contain different types of interfaces such as *Serial interfaces and Ethernet interfaces*. Fig. 2.4 represents a simple network topology with a simplified routing table. Routers are more intelligent than switches because routers can select the best path according to the parameters of the used protocol such as OSPF, EIGRP, RIPv2, and IS-IS. Besides, routers can determine the whole path for a packet that is desired to transfer between the sender and the receiver. Furthermore, routers can achieve all the functions and jobs that can be done by switches [6, 4].



Routing Table

Network ID	Interface	Hop
10.1.0.0	fa0/0	0
10.2.0.0	S0/0	0
10.3.0.0	S0/0	1
10.4.0.0	S0/0	2

Routing Table

Network ID	Interface	Hop
10.2.0.0	s0/0	0
10.3.0.0	S0/1	0
10.4.0.0	S0/1	1
10.1.0.0	S0/0	1

Routing Table

Network ID	Interface	Hop
10.4.0.0	fa0/0	0
10.3.0.0	S0/0	0
10.2.0.0	S0/0	1
10.1.0.0	S0/0	2

CAM Table

Port no.	MAC add.
P1	MAC1
P2	MAC2
P3	MAC3
P4	MAC4

CAM Table

Port no.	MAC add.
P1	MAC5
P2	MAC6
P3	MAC7
P4	MAC8

Fig. 2.4 One domain network topology based on switch and router

2.3 Wired and Wireless Networks

All types of communication networks are either wired or wireless. Classification of wired networks is based on two types according to the manufacturing material of links that connect the network devices, the main ones are copper cables and fiber optic cables. Wired networks are considered the foundation of wireless networks due to the fact that all the access point should be connected to the fronthaul network's devices through a specific cable. Moreover, the fronthaul devices are joined to the backhaul devices which in turn are connected to the backbone or to the Internet networks. Using wired links in fronthaul and backhaul is due to the enormous amount of data that can be transferred by wired links[7]. In general, the wired networks lack of flexibility due to the difficulty of devices' movement, this is why wired networks are fixed and stationary [8, 9].

On the other hand, wireless networks do not need physical links between the end devices and access point that are attached by those end devices. Therefore, wireless networks are extremely flexible and have the advantage of easy spatial mobility within the radio transmission of a network access point to access the Internet. Fig. 2.5 illustrates the relationship between wired and wireless networks.

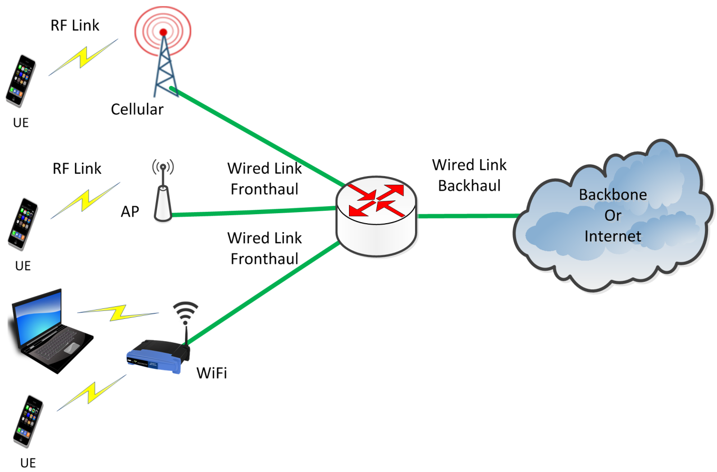


Fig. 2.5 Simplified the relationship between wired and wireless networks

2.4 Network Device Identifiers

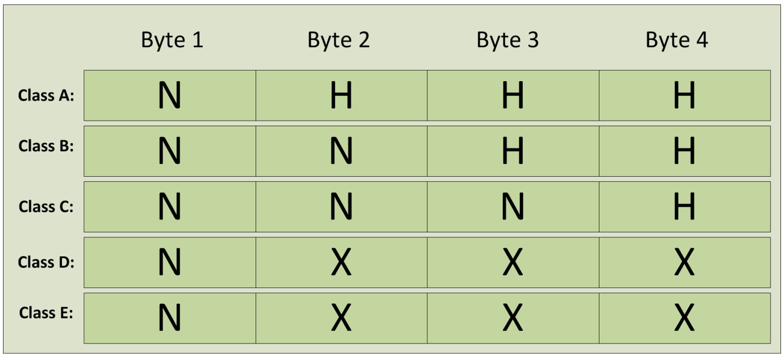
Device identifiers are used to recognize the different devices in a network, which connects to each other. There are several identifiers used by devices such as device name, MAC address, IP address, etc. Any device that has the ability to connect with another device should use at least one of these identifiers. IP (v4 or v6) addressing is the most important of these identifiers [10]. The smartphone is a good example of these devices which use more than one identifier to connect to a base station or to other devices. Since the 5G networks rely entirely on packet switching technology, i.e., these networks use IP and MAC addresses as packet guider in routing and forwarding process to reach its destination.

2.4.1 IP Address

The Internet Protocol (IP) addressing is the most famous ID is used by network devices to introduce itself in communication systems. From an IP address, the location and position of a device can be determined. i.e., two principal objectives are presented by an IP address. Firstly, it identifies a network interface and host interface that attaches the network access. Secondly, it provides information about the location and position of a host in an attached network. As a result of this binding, the network devices can establish a route to that host. The IP address can be split into two major parts, the first one named network prefix ID represents the network ID or topology ID (also known as locator ID) which related with both the network and the subnet. The second part is known as the host ID (also known as the host location identifier), which represents the interface ID of a device or the host identifier [4].

2.4.1.1 IPv4 and IPv6 Addressing

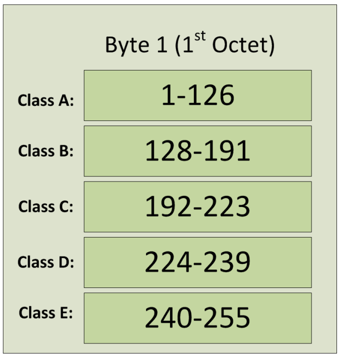
IPv4 (Internet Protocol version 4) is consisted of 32-bit address assigned to the hosts. An IP address fit to one of five network classes (A, B, C, D, or E) shown in Fig. 2.6 and is written as 4 bytes divided by periods, i.e., dotted-decimal format. Each address consists of a Network ID (NID), an optional Subnetwork Identifier (SID), and a Host ID (HID). The NID and SID together are used for routing packets, and the HID is used to address each host individually within a network or subnetwork. The IP directs the packets based on routing protocols from the source network where the packets are



	Byte 1	Byte 2	Byte 3	Byte 4
Class A:	N	H	H	H
Class B:	N	N	H	H
Class C:	N	N	N	H
Class D:	N	X	X	X
Class E:	N	X	X	X

Fig. 2.6 IPv4 network classes

generated to the target network. In order to extract network and subnetwork information from the IP address, a subnet mask is used which is also called an Internet address [11]. All the five classes are identified by the first byte or octet of IP address as shown in Fig. 2.7.

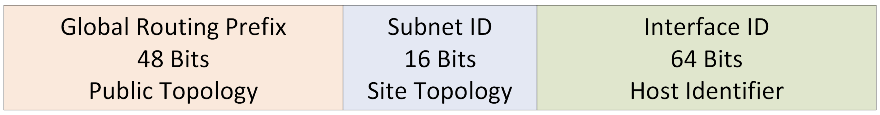


	Byte 1 (1^{st} Octet)
Class A:	1-126
Class B:	128-191
Class C:	192-223
Class D:	224-239
Class E:	240-255

Fig. 2.7 First byte IPv4 network classes identification

The continuously growing of the Internet in a way nobody was expecting the IPv4 addresses will be depleted after 30 years [12]. In spite of all the clever tricks to increase IP addresses such Network Address Translation (NAT) and Variable Length Subnet Mask (VLSM), but those tricks did not solve the problem. Hence, this problem motivated the scientists to find a viable solution to this dilemma. The solution is the IPv6 address, it consists of 128 bits (16 bytes), that give 2^{128} IP addresses. There are no network classes for IPv6 such as in IPv4. IPv6 and IPv4 are not compatible with each other so a new protocols have been replaced or upgraded to be suitable with IPv6 [13].

The structure of IPv6 address is shown in Fig. 2.8. The global routing prefix indicates for the individual site of a certain organization that is connected to IPv6 networks. SID is used to identify subnets of an organization site. Subnets are created by an organization for an efficient routing



Global Routing Prefix 48 Bits Public Topology	Subnet ID 16 Bits Site Topology	Interface ID 64 Bits Host Identifier

Fig. 2.8 The structure of IPv6 address

infrastructure, also for multiple levels of addressing hierarchy. Interface ID (IID) in IPv6 is equivalent to the HID in IPv4. It refers to the interface on a particular subnet inside the site [14]. In other words, IID can be defined as the access port or interface of a router or any device has the ability to access to a network. In all of our proposals, the IPv6 has been suggested to be used as a device address because of its properties and features such as auto-configuration addressing and providing the enormous number of IP addresses.

2.4.1.2 MAC addressing

Every network device should have a unique physical address or what is known as the MAC address (IEEE 802-48bits) which consists of 48 bits [14]. The MAC address is the hardware-level addressing that is used by layer two in the OSI model to communicate devices on the network. To generate the IID which to be used with IPv6 addressing, it should be composed of MAC address and padded hexadecimal number (FFFE) to form IEEE EUI-64. Fig. 2.9 illustrates the modification that must be done on the MAC address to produce a new standard for IID addressing which is represented in IEEE EUI-64 [14, 15]. The term interface has different meaning delimited by the context: In the IP context, it indicates the network attachment port of an IP stack. In the 3GPP context, it involves the entire vertical protocol stack between pair network devices, In the software structure meaning, it indicates the defined communications between two software modules, and seldom it is named for

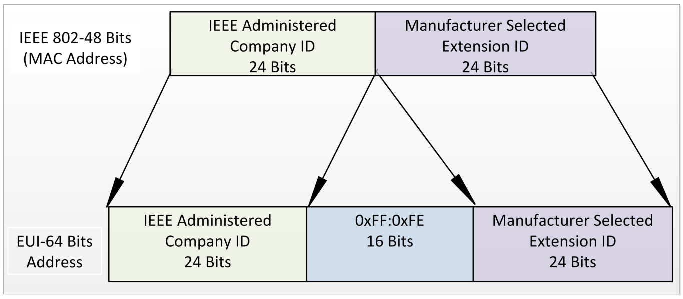


Fig. 2.9 The structure of the EUI-64 address

an Application Programming Interface (API). An IID is the host portion of an IPv6 address. It is typically 64 bits in length. The IID can be associated with a hardware address (MAC) which can be algorithmically or manually generated.

2.5 Mobile Networks

Mobile communications have seen various developmental stages such as First Generation (1G), Second Generation (2G), Third Generation (3G), Fourth Generation (4G) and the next generation or Fifth Generation (5G) is subjected to development and implementation. The brief description of the generations of mobile communication systems is given in the following sections:

2.5.1 1G

For the duration of the early 1980s, analogue cellular mobile systems were used. At that time, each country was developing its own system, control usage within one country margins. This analogue cellular system was suffering from limitations in both data rate capacity, and support for data traffic. The 1G mobile system is considered as a circuit-switch analogue telecommunication scheme [16, 17].

2.5.2 2G

The 2G was globally spread at the beginning of the 1990s. the spectrum frequency used by the Global System for Mobile (GSM) range between (806 to 2890) MHz and the modulation technologies were employed are known Frequency / Time Division Multiple Access (F/TDMA). The throughput was between (270.833 to 473.6) kb/s. The 2G had mixed circuit-switch and packet-switch technologies [17]. Fig. 2.10 illustrates the simplified 2G system architecture.

2.5.3 3G

The third generation was released earlier in the 2000s. The 3G architecture kept the mixing technologies of circuit-switch and packet-switch. The operating frequency range (1885 to 2200) MHz bands have been used by the service providers. The bit rate of the download was 14.0 Mb/s and upload was 5.8 Mb/s [18, 17]. Fig. 2.11 illustrates the simplified 3G system architecture.

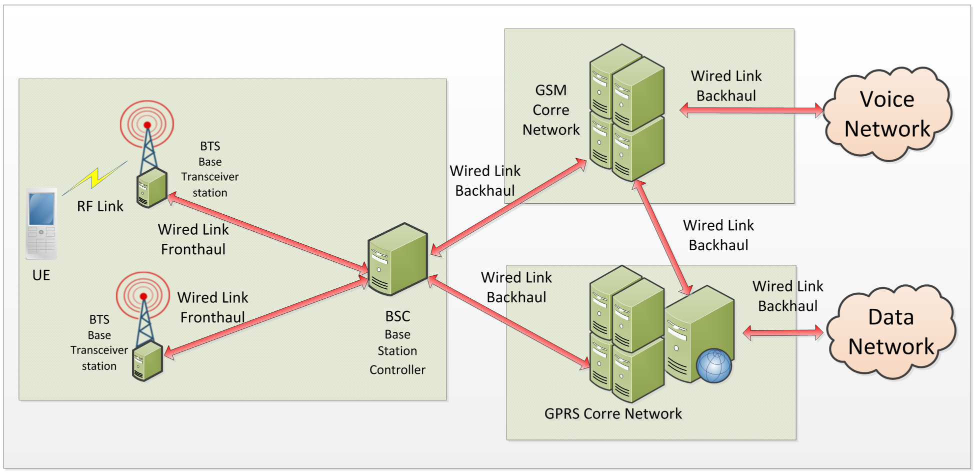


Fig. 2.10 The simplified 2G system architecture

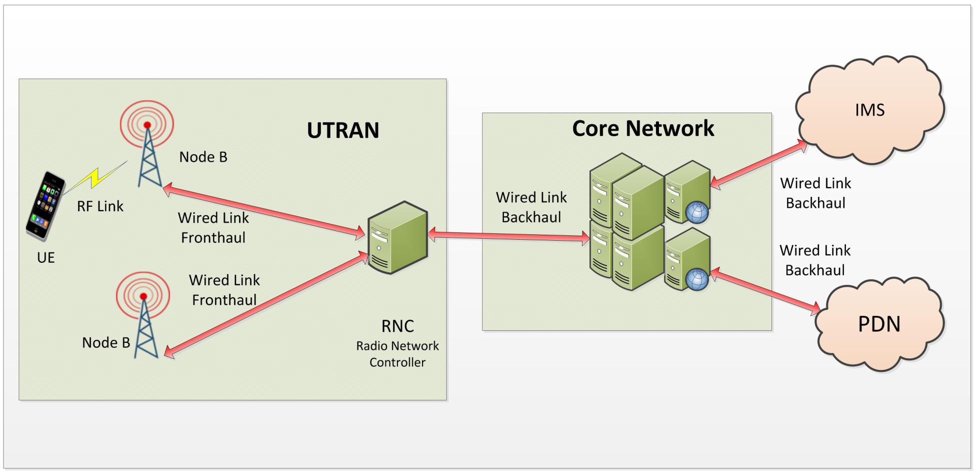


Fig. 2.11 The simplified 3G system architecture

2.5.4 4G

The 4G inherited all the powerful developed technologies of mobile network services to produce ultimate quality services to subscribers. The operating frequency range (700 to 2600) MHz bands have been used by the service providers [19]. The dramatic exponential growth of using mobile data by users, education, business, and many governments and non-government services. Also, the rapid development of applications that require a high quality of service yielded to born 4G [20] . Although, the advanced technology and efficient of 4G which utilizes the extreme range of Bandwidth (BW) in present time. However, it has not been following the burst demands of high-quality services of applications that the costumers desire. The 4G has covered the requirements of its specifications of peak bit rate at 1Gb/s for low mobility and 100 Mb/s for high mobility. The pressing need to satisfy all growing requirements have pushed to planning for a new generation has the ability to deliver a high bit rate larger than previous generations[17, 21, 19]. Fig. 2.12 illustrates the simplified 4G system architecture.

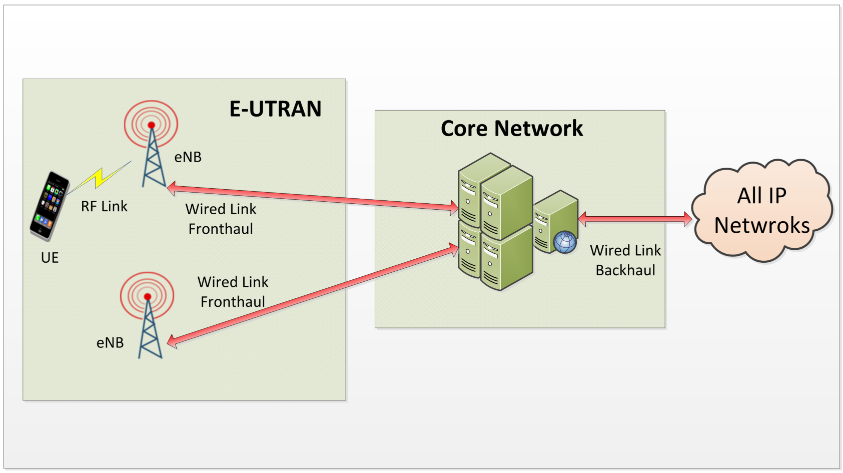


Fig. 2.12 The simplified 4G system architecture

2.5.5 5G

The biggest challenges for the 5G mobile networks is that providing the increasable capacity to accommodate the growth of mobile data traffic. Smartphone's technologies and applications are continuously improving, the number of their users are also increasing, it is logically, this growth in downloading application demands and the number of users leads to surpass mobile network's capability. These obstacles should be tackled by new ideas on communication networks infrastructure to be able to feed the growing demands with efficient performance. In order to achieve these concepts and to get the practical gain of mobile networks advance, which should be able to handle this increase without increasing the cost of those networks [22].

5G mobile networks should be supported by future communication networks that must meet the most significant requirements of 5G mobile networks as listed in Table 2.1. The listed parameters in Table 2.1 are considered by IMT-2020 system targets, which maybe are subjected to more investigations and technological advancement [23]. At the same, these communication networks should decrease CAPEX , OPEX, and power consumption of networks infrastructure cost. Furthermore, the self-organization network functions should present to manage communication networks [24]. 5G communication networks will be more than a new wired or wireless interfaces. The 5G is a comprehensive system architecture incorporates a wide range of items that present virtually a new wired or wireless interfaces which rely on the software infrastructures, such as the virtualizing, slicing, isolating, sharing, and splitting hardware resources by extracting software functionalities of the communications system, furthermore, novel concepts for transmission protocol structures for enhancing the fronthaul and backhaul links performance. All those factors should present to accomplish the highest adaptability to satisfy promised service obligations. Hence, the 5G is not a special technique. It is a design of an extremely manageable, scalable, and flexible communication networks. [3]. Fig. 2.13 illustrates the simplified 5G system architecture.

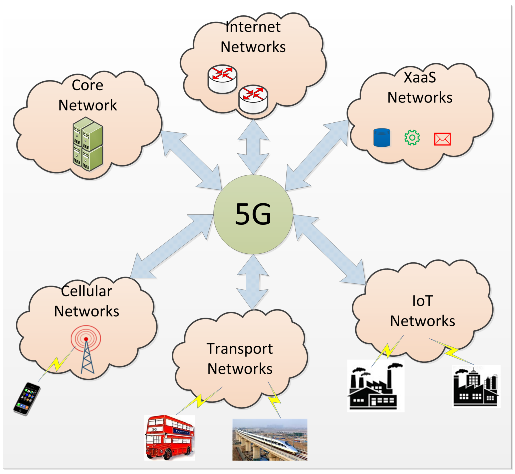


Fig. 2.13 The simplified 5G system architecture

2.6 Evolved Packet System Entities Functionality

The architecture of Evolved Packet System (EPS) consists of two prime parts:

- Long-Term Evolution (LTE) is a standard for fast radio communication for User Equipment (UE), based on the Global System for Mobile communications (GSM)/Enhanced Data rates for GSM Evolution (EDGE) and Universal Mobile Telecommunication System (UMTS)/High Speed Packet Access (HSPA) technologies. LET enhanced the capability and data rate by applying an unconventional radio interface concurrently. It covered the radio interface which

Table 2.1 5G network system targets [23]

Parameter	Target Value	Description
Data Throughput	10 to 20 Gbps	Highest data rate available for each device/user.
End-to-End latency	1 ms or less	Excluding RF registration latency.
Delivered data rate	100 Mbps to 1Gbps	Available data rate to a device/user.
Mobility	Up to 500 km/h	Highest speed at which a device/user can receives data.
Density of connected devices	$10^6/km^2$	Ability to serve connected devices per unit area.
Energy consumption	At least 10 times less	The 5G network should not consume more power with presenting improved features.
Traffic density per unit area	10 $Mb/s/m^2$	Total data transfer rate for each geographic region

uses Orthogonal Frequency Division Multiple Access (OFDMA), and User Equipment (UE). LET defined the bearer Quality of Service (QoS) that is allocated to UE requirements [25].

- The System Architecture Evolution (SAE) has a flat structure, SAE covers the Evolved Packet Core (EPC) which is known as Core Network (CN), EPC included several entities such as Mobility Management Entity (MME), Packet Gateway (PGW), and Serving Gateway (SGW). Besides, some logical entities which are considered as part of the CN, such as Home Subscriber Server (HSS), and Policy and Charging Rules Function (PCRF). Fig. 2.14 shows the key elements that build and form the EPS system.

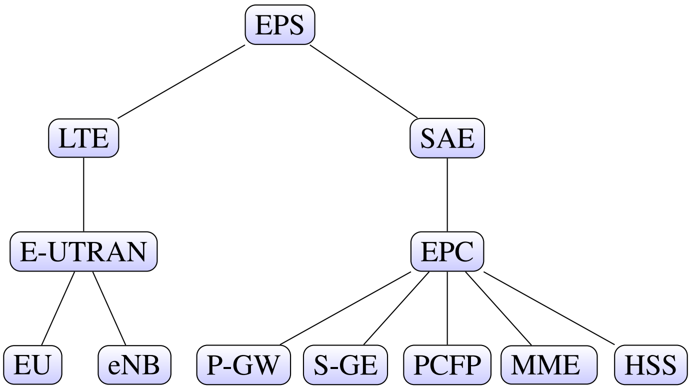


Fig. 2.14 The main elements of EPS structure

2.6.1 Evolved UMTS Terrestrial Radio Access Network

The Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) consists of many eNBs. E-UTRAN handles the EPCs radio communications between the eNB and UEs. It does not have a centralized controller, hence, it is considered as flat architecture.

- eNB can be defined as a wireless access point. It is different from its predecessor of base stations in that eNB is directly connected to network routers. There is no more intermediate controlling devices, such as Base Station Controller (BSC) as in 2G, or Radio Network Controller (RNC) for 3G mobile systems. eNBs connects to the EPC via a Serving Gateway (S-GW). eNBs could be interconnected with each other by enhanced mobility means of the X2 interface link. The protocols between the UE and the eNB are called Access Stratum (AS) protocols [26].
- UE can be as an air interface rival of the eNB and has two components:

  1. Mobile Equipment (ME) involves two components, called the Mobile Terminal (MT), which is utilized for the air interface for all communication functions and Terminal Equipment (TE) where the data streams are terminated.
  2. UMTS Subscriber Identity Module (USIM) is a smart card that holds the user identity, home network identity, stores authentication and encryption keys, and performs the user authentication [27].

2.6.2 Evolved Packet Core (EPC)

The EPC consists of many network devices that are routed and manipulated the arrived packet to deliver it to its target destination. The following subsections briefly illustrate the most important devices (entities) that involve in an EPC of mobile networks:

2.6.2.1 Mobility Management Entity (MME)

The MME is responsible for high-level signaling of the UE, such as the issues of authentication and security. Additionally, it manages the continuity of data streams that unrelated to radio communications.

Every subscriber is assigned and overseen by a specific MME that looks after a distinct geographical area, but if the user moved far away, it would be served by another MME to look after it.

2.6.2.2 Policy Control and Charging Rules Function (PCRF)

The PCRF detects the service flows, and it is responsible for charging policy. It delivers based on users subscription, the Quality of Service (QoS), and the authorization that identifies how a data stream will be moved through the network.

2.6.2.3 Packet Data Network Gateway (PDN-GW)

The PDN-GW, also known as P-GW, It is responsible for allocating the IP addresses to the UEs, perform QoS application and per flow-based charging that follows rules of the PCRF entity.

2.6.2.4 Home Subscriber Servers (HSS)

The HSS entity holds the databases of the subscribers public and private identities, security variables of users, and the position and location information functions. It contains (i) Home Location Register (HLR) which is in charge of storing and updating all the users' subscription information. (ii) The Authentication Centre (AuC) which is responsible for generating security information from users' identity keys. The security information is shared with neighbor networks terminal authentication, also is used for radio path ciphering integrity protection

2.6.2.5 Serving Gateway (S-GW)

It has crucial functions to empower users to connect the EPS network. S-GW creates, deletes, and modifies bearers for each UE that connected to the EPS, performing of these functions depend on per Packet Data Network (PDN) that established connections for every UE.

2.7 Mobile IP (MIP)

The main utilization of IP address is to deliver data packets between nodes on the communication networks and the Internet. These nodes should have a unique IP address to allow forwarding packets across different networks on the Internet. This scenario is somewhat easy for fixed nodes. Another scenario shows more difficult when a node moves between different networks. To solve this dilemma, a Mobile IP (MIP) emerged as an extension to the IPv4. The MIP allows a node to move amongst different networks [28]. However, in the IPv6, mobility supporting is directly built into the protocol, when the MN connected to its service provider network which is known as a Home Agent (HA), the IP address which is given to that MN known as a Home Address (HoA). The HoA is a permanent and unique IP address assigned by the HA to an MN belongs to it. When the MN moves and connects to other networks, this network is called a Foreign Agent (FA). A new IP address is assigned by the FA to this MN, this a new IP address is called Care of Address (CoA) [29].

2.7.1 MIP Components

The MIP is designed to allow mobile devices to move from one network to another while maintaining a permanent IP address. It enables a node uses an IP to retain the same IP address and maintain existing communications while moving from one network to another. The MIP concept consists of four basic elements they are:

1. Mobile Node (MN), it might be a cell phone, laptop, or any personal gadget uses wireless communications and network roaming capabilities.

2. Home Agent (HA), it is a router on the home network serves as the anchor point for communication with the MN. Also, it is responsible for forwarding and receiving packets from/to the MN. It uses a technique known as tunneling to achieve a connection establishing between the HA and a reachable point for the MN in the foreign network.

3. Foreign Agent (FA), it is a router that may function as the point of attachment for the MN when it moves to a foreign network, It receives packets from the HA and delivers them to the MN which is hosted by the foreign network.

4. Sender Node (SN), it represents the destination or end host (fixed or mobile), which may be a browsing server or another mobile node.

The procedure of the MIP has three main phases:

- Agent Discovery: The mobility agents advertise their accessibility by broadcasting means. This procedure involves broadcast of advertisements by the routers to their connected sub-networks. An MN receives configuration instructions from its foreign and home agents during the transmission of the amended advertisements information.
- Registration: When an MN visits an FA, it initiates the process of the registration. It sends its home network address and home agent's address. The MN will get a CoA from the FA, and it will contact its HA to register its CoA. This process is called binding. The HA redirects the packets traffic to the MN [29].
- Tunneling: After the MN registered its CoA to the HA, it will receive all packets from its HA through tunneling the traffic to the CoA of the MN. [30].

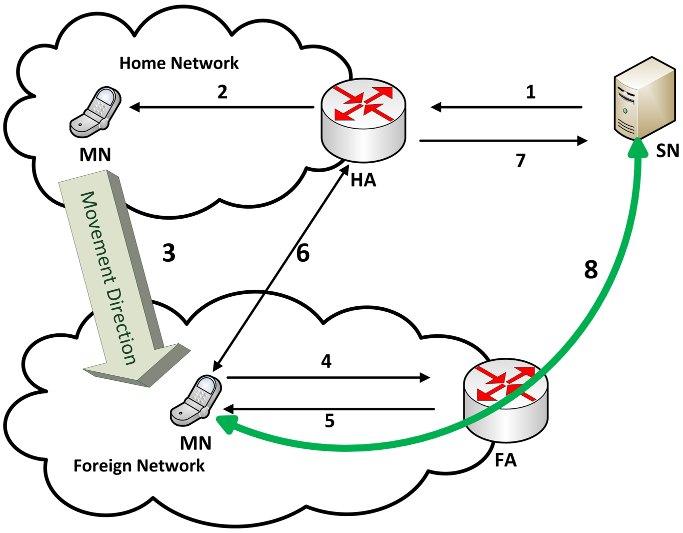


Fig. 2.15 Simplified MIP architecture

2.7.2 Basic MIP Process

MIP process is the means for offering seamless roaming to mobile devices on communications and Internet networks. MIP can be considered as obtaining a new IP address from the visited network. Three major mechanisms cooperate to achieve the MIP process. First, the discovery mechanism defines that UE can determine the new attachment point (request a new IP address). Second, once the UE obtains an IP address from the visited network, it registers with the home agent. Lastly, MIP defines a mechanism to transfer data packets to the UE when it is away from its home agent [29]. Fig. 2.15 illustrates the essential steps of the MIP procedure, according to the reference [31].

1. Step 1: The SN sends packets to the HA of the MN, the SN knows the HoA of the home network of the MN.

2. Step 2: When the MN at home network, it uses its permanent HoA, all packets are directly forwarded by the HA to the MN.

3. Step 3: The MN moves toward the foreign network, and it enters the coverage range of the foreign network domain.

4. Step 4: The MN sends a request to the FA for acquiring a temporary IP address and registration itself at the FA.

5. Step 5: The MN acquires a temporary IP address from the FA, this IP address known as a CoA.

6. Step 6: The MN sends this CoA to its HA for registration and binding, through a secured connection called the binding update. The HA binds the CoA and the permanent home address of the MN. The HA will deliver all the data to the MN by using the CoA.

7. Step 7: The SN is not aware of changing of the MN's IP address. The HA manages and handles of the MN's mobility and sends the binding update (permanent and CoA of the MN).

8. Step 8: The SN sends the data to the MN by using CoA of the MN, without passing these data across the HA. This is achieved by utilizing route optimization.

This process is clearly facing the handover delay problem during user mobility among networks. This will significantly affect end user's experience. This delay is caused by acquiring CoA, registration CoA at FA and HA, binding update at HA and SN. The handover delay can obviously notice in real-time applications such as voice calling. The MN be unreachable for a period of time due to the handover mechanism in the MIPv6 [28].

2.8 C-RAN

The concept of the C-RAN is structured based on gathering baseband processing units in a single place to serve many of far transmitting antennae which are known as Remote Radio Heads (RRHs). The idea of C-RAN has been submitted by several mobile network operators such as Telefonica, SoftBank/Sprint, France -Telecom/Orange, and China Mobile as well as industrial companies of mobile equipment vendors such as Nokia-Siemens, Alcatel-Lucent, Light Radio, and many other companies. The C-RAN emerged to address the lacks of spectrum efficiency, flexibility to deploy further RRHs, scalability, and energy consumption. Hence, C-RAN concept will support the next generation of mobile networks at least from the views of the power consumption and spectrum resources [32, 33]. The structure of the C-RAN consists of several RRHs that are served by a Baseband Unit (BBU) pool. The distributed RRHs consume very low power and are joined either by wireless links or optical fibers links to the BBU pool [34]. Fig. 2.16 shows the C-RAN paradigm system.

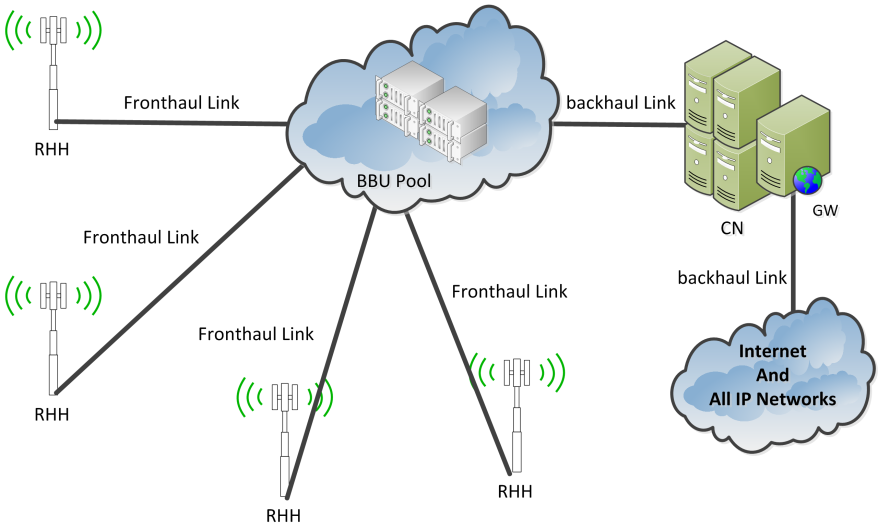


Fig. 2.16 Simplified C-RAN system architecture

2.9 Virtualization

The Virtualization is known as the procedure of creating virtual forms of physical resources to simulate the equivalent characteristics of the physical resources. In the computer world, the virtualization term is used to refer to the abstraction of fundamental computing resources of a physical device to emulate the work of that physical device features or other elements of a physical device to accomplish a task logically (i.e., by software programs). The created Virtual Machine (VM) [35] has a virtual processing unit, a virtual storage unit, a virtual operating system, and application programs which are used to achieve specific tasks by partitioning and sharing a physical resource amongst one or several VMs. Fig. 2.17 illustrates the basic architecture of created VMs within a physical machine. One of the most significant yields of the virtualization technology is the Cloud Computing (CC) [36] which runs the application services away from the people who use computers, the CC has been utilized as the platform of choice for a lot of online services and network services. By using virtualization, the hardware resources and software application services are flexible and not firmly joined by particular hardware or software.

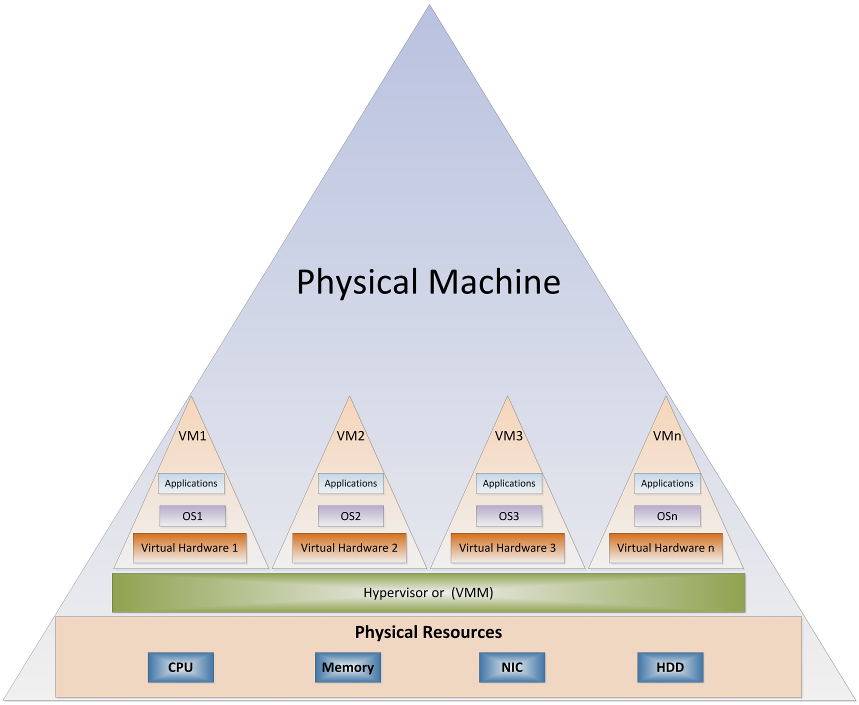


Fig. 2.17 Correlation between the physical and virtual machines

Another benefit of the virtualization is the Network Functions Virtualization (NFV) [17] which draws a significant theory of transmutation for mobile and Internet service provider's networks migration from hardware to a virtual infrastructure, motivated by the goals for reducing CAPEX and OPEX, providing personalized services, and increasing flexibility. Furthermore, virtualization empowers the

Software-Defined Networking (SDN) as a solution within dynamic network circumstances [17], new protocols and standards need to be programmed and adapted by clients. Hence, several organizations try to standardize open networks and protocols such as the Open Networking Foundation (ONF) and OpenFlow protocols have emerged to assist in the deployment and implementation of SDN-enabled systems [37–39]. As a result of virtualizing of real servers, the degradation in performance of virtualized network resources due to sharing the same physical resources, this degradation of resources can be substituted by advanced technologies of high-performance physical network hardware industries.

2.9.1 Network Virtualization

Virtualization of the network has emerged to eliminate the constraints posed by physical network devices, as well as allowing for the flexible provision of network services. Furthermore, many virtualized networks can be created that work through a real network infrastructure. In the network virtualization, the virtualized allocated resources of a virtualized network have been controlled and monitored by the hypervisor, such as resources are Network Interface Card (NIC), buffer size, link capacity, and ports feature of routers or switches to individual virtualized network resources (slices) [40]. Network virtualization technology has become more efficient and worthy of implementation in a virtualized network environment due to the development in industrial technologies to produce hardware for the servers able to manipulate a large size of data parallelly at the same time [41].

2.9.2 Network Slicing

5G network promise to improve the performance of different serving communication networks such as mobile networks, health networks, industrial networks, autonomous car networks, and transportation networks, these diverse kinds of networks need very low latency, very high data rate delivery, and real-time auto-management. Since 5G networks are capable of fitting these requirements, they will be suitable to serve these networks in efficient, reliable, scalable, and flexible infrastructure [17]. In order to achieve these requirements for designing and building efficient and reliable networks, a new technique should be used to assign a specific resource to be part of a network or several networks. This resource can be exploited to achieve a specific task as need and can be shared by more than one server. This concept is known as a network slicing [42].

The network slicing principle allows a high dynamic and coordination for network functions deployment established on various network service conditions. In addition, the network slicing technology provides the concept of every slice is designated to carry out a set of essential network functions based on different service conditions. Also, to create flexible potential methods of distributing network functions on demand and need. Network slicing has emerged as a novel structure that has an ability of flexible sharing of the hardware or software of physical or virtual network resources, such as functions, servers, or network entities to perform a specific task. In order to achieve network slicing of resources and functions, these functions and resources should be used to share and cooperate of these resources and functions among virtual machines or real machines, whether those machines are on a single network or several networks [38, 43, 44]. Fig. 2.18 illustrates the concept of network slicing resources.

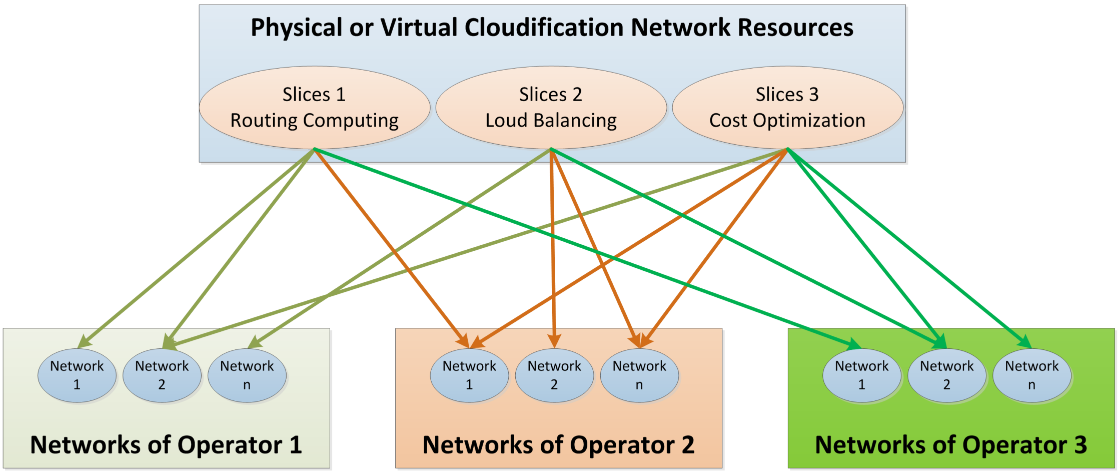


Fig. 2.18 Network slicing concept; sharing and partitioning the virtual or physical network resources

In brief, network slicing refers to a set of physical or virtualized network resources that are isolated for implementing a particular job on networks. Each slice could independently be shared and conducted network resources. Also, it could include computation and network resources that allow a network operator to change the behavior of forwarding, routing, and computation not only for one network but also for other operator's networks.

2.9.3 Slicing Versus Virtualization

The techniques of slicing and virtualization are different in the environment of communication system resources. Slicing is the process of allocating specific resource to be part of a network slice. It can be considered as a shared resource. In other words, it is not necessarily to virtualize and share the sliced resources and in this case, the slicing is known as resources-partitioning. While the virtualization of resource means that the same physical resource can be used and shared by multiple slices [45]. Cloud computing provides services from large, highly virtualized or real servers using application and pooled resources [46]. Three approaches are defined by the National Institute of Standards and Technology (NIST).

- PaaS: Platform as a Service which provides a high-level integrated environment to build, test the prototype, and set up custom applications.
- SaaS: Software as a Service which supplies special-purpose software that can be remotely accessed by users through the internet network.
- IaaS: Infrastructure as a Service combines hardware, software, and tools to reach software application environments [44].
- AaaS: Anything as a Service (or XaaS) it can be defined as a virtualization and an abstraction the soul of any function or job that is achieved by hardware, software, physically, or logically in order to put as a service based on CC [47–49].

Sharing of the physical hardware resources in cloud services are manipulated in a virtualized environment to be rendered by forms of abstract slices to the VMs. Hence, the Internet and mobile communication services can be virtualized and provided by a set of VMs in a virtual environment to fulfill the functions and jobs of the physical servers [50].

2.10 Software Defined Networking (SDN)

Traditional networks can be classified into three layers, known management, control, and forwarding planes. The management layer renders services to monitor and configure the network. The control layer forms the data needed to build forwarding and routing tables, which in turn are utilized by the forwarding layer to conduct data packets to ingress and egress ports. In conventional network models, both the control and the forwarding layers are tightly built within a single network device (e.g., routers and switches). This model provides efficient performance. However, when the size and complexity of the networks scaled, a need to develop a new structure has appeared. The essence of SDN design is the separation of control and forwarding layers. The network devices level become merely forwarding devices, and the network control level is migrated to an independent item called a network controller [51].

The SDN concept depends on separating the controlling packets from data packets to be processed and forwarded by different network devices to reach the data to their destination. The controlling devices work as commander and manager devices which administer the forwarding devices, which are being as only forwarding data devices in the network. SDN concept has emerged to overcome the different types of network devices that have been producing by many different companies. SDN has been developed and standardized by an Open Networking Foundation (ONF), which is a non-profit consortium. ONF defined the SDN architecture as a separation of the control plane and data plane. The intellect of the network is logically centralized, and the implicit network devices are extracted away from the applications [52]. The idea of SDN based on four pillars, they are:

- Separation of the CP from the DP whether in physical or virtualized networking resources.
- Forwarding decisions are made by SDNc which is centralized away from the forwarding data plane devices.
- Packet forwarding based on the flow rather than the destination IP address.
- Programmable software of the network functions interacts with the forwarding data plane devices via Application Program Interface (API) under the management of the SDNc.

In general, SDN concept architecture composed of four principal elements as shown in Fig 2.19. Each of these elements is explained briefly below [53, 54].

1. Applications layer involves the programmable network applications that exchange the information with the SDNc, which interacts with the extracted information that is constructed by the

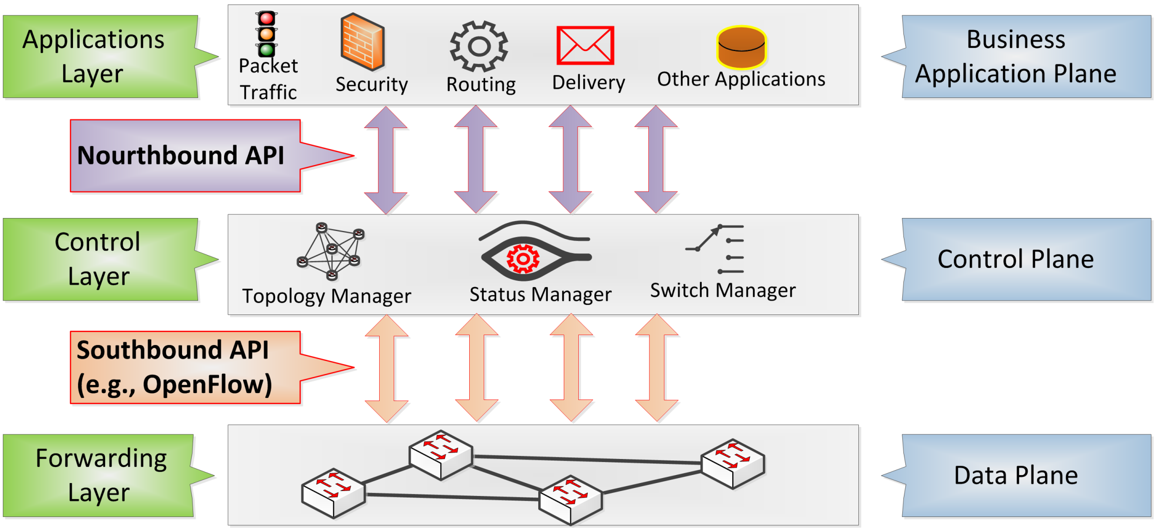


Fig. 2.19 SDN concept architecture.

application layer about the network infrastructure and status. It introduces programmability, which is a key concept in the SDN network. Network programmability affords possibilities for network innovation with a tremendous number of network applications such as monitoring, security, packets' traffic, packet routing, and other applications.

2. Control layer contains the main element of the SDN model, (i.e., SDNc) which manages and makes all the forwarding rules and decisions based on its a global view of the network. The SDNc is the creative element that is responsible for directing and making decisions regarding the data flows that enter the underlying SDN network devices through northbound and southbound APIs. Centralized view enables the network administrator to control easily of monitoring and management a whole network. Additionally, it reduces mistakes in configuring and expanding network policies. The centralization enhances flexibility; for example, a group of network devices from multiple vendors can be used and abstracted in a single network.

3. In universal terms, API is a group of defined rules for communicating between various software parts. The API is a collection of routines, protocols, and tools for creating software applications to allow different devices' ports to communicate. Generally, an API specifies how the user's programmed applications should interact with interfaces. Three types of API work with SDN concept.

(a) Northbound API connects the control layer with the applications layer. It communicates between the network management station running its network applications and the SDNc.

(b) Southbound API helps efficient control of the network and permits the SDNc dynamically to make modifications based on real-time demands and needs. It connects the SDNc with the real infrastructure devices of the network.

(c) East-West APIs that define the communication of different controllers in the same domain or adjacent domains to interact with each other.

4. Forwarding layer represents the physical devices that forward the data packets according to the rules and actions that are sent by an SDNc via southbound APIs to govern the data flows of forwarding devices. The forwarding plane contains network devices such as routers, switches, and middleboxes, which do not have their controlling logic.

Various opportunities have been conducting by SDN networks to the CC and the virtual networking environments. The key innovation of the SDN concept is the separation of the forwarding data plane from the network management control plane. For instance, the network data flows can be forwarded and directed by non-standard routing protocols. Also, using innovative tools and identifiers for data packets to be handled by the SDN networks.

2.11 Network Functions Virtualization (NFV)

The simplest definition of the NFV is that virtualizing network services to be implemented by physical or virtual machines. Additionally, it is regarded as the software infrastructure of the networks. NFV grew and became achievable with the beginning of using the SDN by networks infrastructure owners. NFV is the reaction of network service operators to their shortage of agility, flexibility, and constant requirement for reliable software infrastructures. It is capable of adequate deployment for on-demand networking and user services. In other words, the virtualized functions of a network can be driven by dynamically covering for the network as desired of the network operator to serve any user or application with SDN support. Furthermore, NFV abstracts functionalities of the physical hardware of the network devices that perform those functions in which to be executed by individual devices, users, or rented virtual network by subscribers belonging to NFV services provider's network. That means the functions are achieved by a form of software stack or plain software situation [51, 55, 56].

The concept of NFV techniques relates with the cloudification idea for many reasons, such as reducing the networks maintenance cost, flexible sharing of the software infrastructure resources, providing multi-tenancy of networking functions, and providing a meaningful increase in energy efficiency due to execute the functions as software instead of using dedicated hardware to fulfill these functions. Additionally to facilitating the utility of the term off-the-shelf servers to run particular functions' software. SDNs are one of the main enablers for NFV. For instance, the network devices (router and switch) can be regarded as an SDN-enabled virtual infrastructure where NFV and application services are used nearby to the position where they are certainly going to be used, which will lead to cheaper, agile and more flexible operations.

In sum up NFV abstracts network functions, empowering them to be established, controlled, managed, and formed by software running on compute nodes, devices, networks whether those are physical or virtual. NFV consolidates virtualization and cloud technologies to accelerate the fast development of new network services with flexible scale and virtual computerization. These technologies are closely correlated as NFV and SDN.

2.12 SDN and Network Devices

Traditional router decapsulates the received packet to verify the destination IP address of the arrived packet when it processes the forwarding decisions to that packet. Depending on the routing tables information that is saved by the router, the forwarding decisions are made to direct the data to its end. Based on the matching of the destination IP address with the routing table records, the packet will be passed by an appropriate outgoing interface to reach the destination. This procedure is repeated for every packet and in all routers to determine the best path for a packet between the source and destination for each transmission. Thus, if a connection contains a large number of packets, the arrived packets to an interface will have to wait for their role to be processed by the router to forward to their target. This processing and queuing delays are the real causes of the packet loss during the delivery of the data of any live stream. In addition to the queuing and processing, the registration process also leads to losing the data while the roaming of a moving node between the wireless access points.

As mentioned above, the conventional router needs to test every single packet to determine through which network interface this packet should be forwarded. In contrast, the SDN network only the first packet of a flow is subject to checking the destination IP address, and all the other packets follow the first packet as a flow of a data stream, this approach decreases the processing time, processing power consumption and queuing delay at the ingress and egress interfaces of the network devices. SDN provides flexible tools and mechanisms to make mobile communication systems (fixed or stationary part) more reliable, adaptable, and manageable. It makes the network programmable by separating the control plane from the data plane. This decoupling allows the system to know what goes where and sends the packet to whom. SDN relies on switches that can be controlled and programmed through an SDN controller using an OpenFlow protocol to connect the controller to the underneath forwarding devices which also offers to separate services from the underlying physical network devices and enables an entirely new approach to build networks architecture.

In summary, the ordinary router needs to test each packet to decide to which network that packet should be routed to its endpoint. While in SDN and it's complementary OpenFlow environments only the first packet is subjected to check any field of the packet's header to forward all packets belonging to a single flow to their destination. This procedure minimizes the processing and queuing delay times at the ingress and egress interfaces of network devices, thereby reducing data loss, especially in case of live data streams.

2.13 OpenFlow Protocol

OpenFlow is a predominant control protocol that connects the control layer with the forwarding layer in SDN-based network. It can be considered as a tool for exchanging control messages between an SDN controller and physical devices of data forwarding layer. In 2008, the first practical SDN controller used OpenFlow protocol to transfer control information between the control layer and data forwarding layer [57]. The SDN controller uses the OpenFlow protocol (Southbound API) for communicating and configuring the DP devices (switches, routers, etc.) to send data packets on the best path. The OpenFlow is the first standard programmable protocol for network communication interfaces is designed for SDN. This protocol depends on the separation of the CP from the DP. Also, the OpenFlow protocol manages a part of DP roles and actions, but it does not specify the SDN controller behavior. Several releases have existed of OpenFlow protocol specifications; the latest version is more powerful, due to its feature sets support IPv6, metering, rerouting, policing, Multiprotocol Label Switching (MPLS), and more scalable control. The first release named Version 1.0.0, which launched in December 2009, despite many points prereleases appeared before then that were made ready for test purposes as the specification developed. The OpenFlow protocol has evolved significantly with each version of OpenFlow release. The current version of OpenFlow is 1.5.1 (until writing this thesis). However, version 1.6 has been available since September 2016, but accessible only to ONF's members [58].

As previously explained, the SDN can be applied to both physical and virtual networks as well. Consequently, OpenFlow protocol also can be applied in real networking switches as well as virtual networking environments. The exchanged information between the controller and the underlying OpenFlow devices (switches) contains *Rules*, *Actions*, and *Statistics (Stats)*, these are principal fields which form the *Flow Table Entry*. The *Rule* field consists of sub-fields such as switch port number, source MAC and IP addresses, destination MAC and IP addresses, protocol type, virtual LAN ID, port IP address, sources and destination transport numbers, etc. While the *Action* field involve the action should be applied by the OpenFlow switch on the received packet. Such as actions are applied for packet forwarding to a specific port or several ports, drop the packet, send to a traditional processing pipeline, and forward the packet. The last field provides the number of packets and the bytes through the packets and byte counters. Fig. 2.20 illustrates the entries and fields of the flow table that is created by the SDN controller and that arriving packets have to be subject to them.

In brief, the OpenFlow protocol specifies the interaction between an OpenFlow switch and an SDN controller. The protocol involves a group of information that is carried by the SDN controller to the OpenFlow switch and a set of corresponding messages that are transferred in the reverse direction. These messages enable the controller to program the switch. Depending on the exchanged messages, the controller can define, modify, and delete flows of a table. We can define the flow as a group of packets are conveyed by one network endpoint to another endpoint. The endpoints may be identified

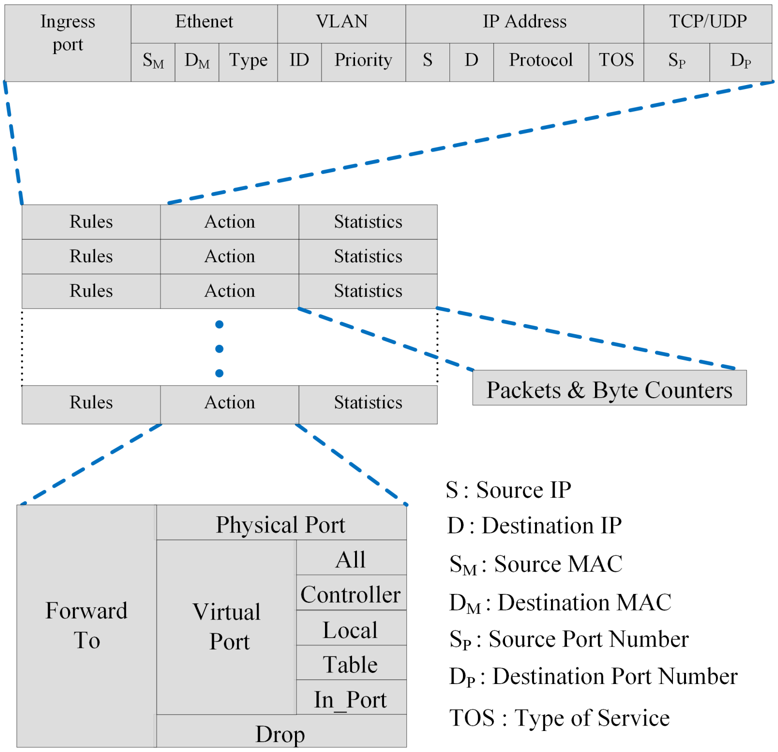


Fig. 2.20 Entries and fields of generated OpenFlow Table

by an IP address, VLAN ID, TCP/UDP port number, layer 3 tunneling tag, or IID amongst other things. The rule field dictates the switch to which constraint should match the incoming packet, the action field in the flow table describes what should the device take action for all packets of a flow. One of the three basic possibilities could the switch treat the incoming packet; forwarding the packet out one or more output ports, passing the packet to the SDN controller for exception handling, or dropping the packet. Fig. 2.21 show exploitation of flow tables in a switch, router, or chipset.

2.14 Smart Virtualization

Network virtualization and network device softwarization techniques conduce to complement each other such as network virtualization, NFV, and SDN. These combined technologies and concepts are used to interact with each other to yield what is called *Smart Virtualization*. Fig. 2.22 shows the idea of smart virtualization according to combining and interacting with different networking technologies

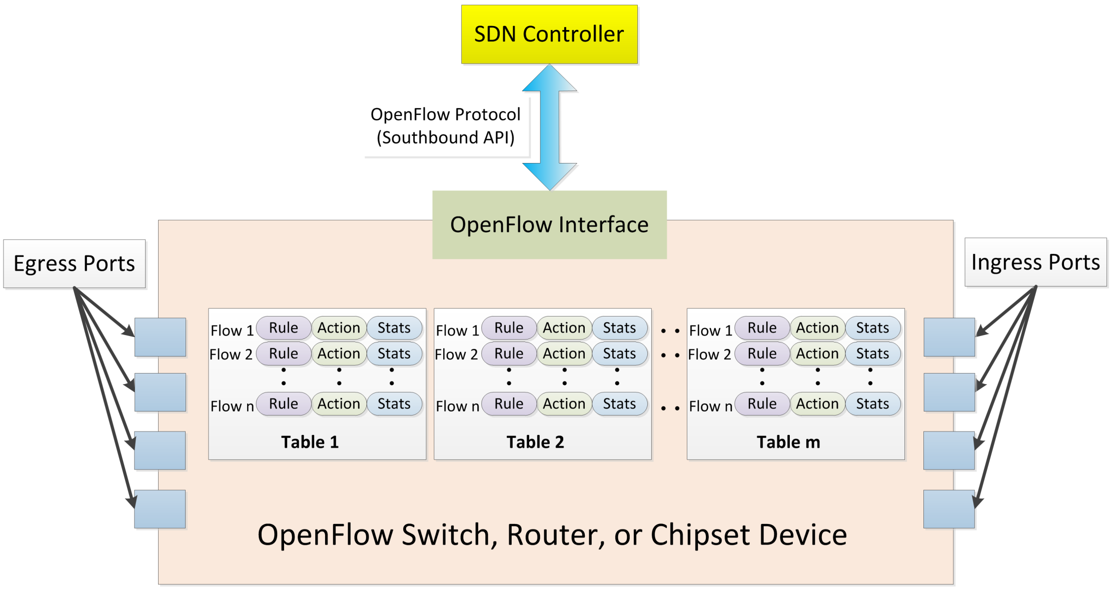


Fig. 2.21 Installed flow tables in an OpenFlow device

to create a smart network. The relationship between the terms that are mentioned above is illustrated by the following paragraphs which give a quick explanation for each one.

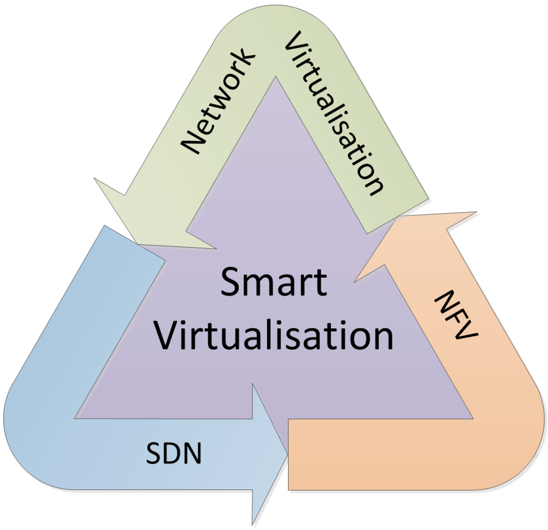


Fig. 2.22 Smart Virtualization.

Network virtualization indicates to abstracting and separating of the network topology from the physical network infrastructure to build it in a virtual environment. This technology allows creating several virtual networks to be deployed by a single physical device or more. Each one of the virtualized networks becomes quite manageable and flexible topology than that of the real network topology.

Network virtualization allows network administrators to design and build networks without requiring to make changes in underlying network infrastructure. The network virtualization idea based on the separation of logical topology from physical topology to mimic the physical network infrastructure in a way different from that idea of SDN which based on the isolation control plane from the data plane of networking infrastructure whether it is physical or virtual.In fact, virtual network technology and SDN technology do not depend on each other's work. The SDN does not necessarily represent the virtual network, and it is certainly not a requirement to achieve the virtual network. The reverse is true, but a virtual network solution can be deployed on an SDN network, as well as SDN can be deployed in a virtual environment.

Since emerging SDN technology powerfully supports network virtualization, which in turn is regarded as one of the most significant networking technique effectively enhances SDN performance. As a result, a flexible virtualized network architecture can be managed by supporting SDN as an activator for network virtualization. That means, the concept of network virtualization can be considered as a directed solution to a specific problem of available devices in network infrastructure, while the SDN technique is a means to manage and control a real or virtual network in an agile and scalable way. In short, network virtualization is independent of SDN, but it could become a fabulous innovated technology with enabling SDN [59, 60].

To understand the close relationship between NFV and SDN, it requires to give a clear definition of NFV. The concept of decoupling functions of a network from underlying physical hardware is as well as decoupling CP from DP concept that is adopted by SDN. In the NFV concept, network functions are provided by virtual devices instead of being tightly associated with dedicated physical devices. NFV enables network functionalities in a flexible and effective transport over networks infrastructure. Although NFV can be achieved by physical or virtual servers without SDN assistance, using SDN with NFV can speed up NFV deployment by providing flexible management and scalable virtualized network infrastructure. By introducing virtualized network functions, these functions can be run on industry-standard servers, routers, storage devices, and switches, rather than using the exclusive task of network devices. This strategy decreases OPEX and CAPEX due to network operators no longer required to spend on costly proprietary physical network devices. Furthermore, network administration is becoming more flexible as NFV services can be modified or introduced quickly to address changing requirements [61, 62].

2.15 Chapter Summary

The trend of communication system architectures goes toward the functionalities of the softwarization services of communication networks. This trend is empowering to emerging technologies such as network virtualization, SDN networks, NFV, and network slicing to be applied by advanced communication network architectures. Smart virtualization refers to the communication networks that utilize those technologies through superior cooperated methods support by novel ideas. Performance of the communication systems depends on several parameters such as End-to-End control delay, the packet forwarding mechanism, and speed of moving mobile nodes. Mentioned technologies could strongly enhance those parameters to optimal values by serving each other with support creative ideas to design efficient communication systems based on *Smart Virtualization*.

Chapter 3

Mini Base Station: Smart Virtual eNB

3.1 Introduction

As it was discussed in Section 2.5.5, 5G communication network systems are geared towards minimizing the size, cost, power consumption, and deployment of physical hardware. In addition, it will increase the transmission data rate, speeds up of the delivery of data, increases performance, and virtualizes physical resources of the support of the Software Defined Networking (SDN) and Network Functions Virtualization (NFV). The most significant challenges for 5G mobile communication networks are providing sufficient capacity to accommodate the growth of mobile data traffic and decreasing End-to-End latency. The technologies of movable smart gadgets have been advancing. At the same time, the numbers of users have been expanding, and intuitively, will lead to exceeding of the capabilities of mobile communication networks. In order to address such obstacles, these networks should be efficient to satisfy expandable needs of mobile networks of effective performance. To realize these goals, mobile networks should be able to handle this expansion without increasing the cost of these networks [22, 63]. Most of the current or proposed mobile communication networks working under 4G conditions have implemented virtualization in both hardware and software infrastructure. The theory of SDN depends on splitting the Control Plane (CP) from the Data Plane (DP) of network devices [64]. This separation is to simplify the control and management of networks. It is also to facilitate progression and development by abstracting the network control functions into a virtually centralized CP that allows for the network management to be programmable by software developers in a smooth and easy way [65]. Making decisions is based on software controllers that are logically centralized away from the network devices that forward the data packets. The forwarding devices can be controlled and programmed through an open interface by using the OpenFlow protocol to become pure data packet forwarding devices [66, 67]. SDN plays the influential role in reducing the power, minimizing the number of physical devices, decreasing the cost of installing a network, reducing control delay time, and in allowing the network administrator to configure the instructions for the CP which are executed by the centralized controller [68]. The integrity of SDN is NFV technology [69], which enables building an End-to-End network underpinned by utilizing virtualization technology standards to empower the consolidation of different communication network devices [56]. Cloud computing paved the way to coping conveniently with mobile communication network infrastructure [70]. The infrastructure of communication systems, whether it be wireless, wired, motionless or mobile networks, the virtualization technology is growing dramatically to design and build these networks. Such virtualization can be seen in mobile cellular air networks, such as Cloud Radio Access Network (C-RAN), which is emerged as one of the results of cloud computing. In C-RAN the BBUs (whether virtualized or not) are gathered in a centralized BBU pool, where the fundamental RF functions are accomplished, such as establishing the initial connection, processing RF signals and managing and maintaining the established connections. The proposed SVeNB has the ability to perform these responsibilities depending on the users' profiles that are received from the BBU and the CN. These user profiles have already processed by the BBU and the CN and sent to SVeNB which covers the local users within its coverage area. This information is stored as a backup,for instance, to be used by SVeNB to empower the resident subscribers who are covered by it, without communicating with the BBU pool and the CN. This process enormously mitigates the load on the CN, the BBU pool, while also lowering decisions making latency, which leads to expediting up the establishment of the connection of the End-to-End device. Our proposal presents a new mini-base station for 5G and beyond for mobile communication networks that use virtualization techniques. The main contributions of this research are follows:

- Proposing a novel mini-base station that utilizes different virtualization technologies to simulate the kernel functions of mobile communication networks;

- Mitigating the burden on the BBU pool and the CN by making connection decisions between users without needing to contact the BBU pool and the CN;
- Decreasing the End-to-End delay by reducing the number of control messages amongst the serving entities of mobile networks.

These advantages have been got by bringing forward the virtualised functionalities of serving entities near end users. Furthermore, the NFV and possibility of abstraction the vital functions from the BBU pool and the CN have helped to obtain those benefits. Besides functions slicing and virtualising, SVeNB's VMs have been configured to simulate these functions slice to perform and complete decisions of the connection session. MATLAB software has been used to simulate and adapt the extracted and simplified the kernel of these functions in order to virtualise and install on SVeNB's VMs to serve the local users that are covered by SVeNB.

3.2 Background and Related Work

There has been much research on virtualizing wireless mobile Base Stations (BSs) and some of this literature is considered here in chronological order. In 2010, Zhu et al. [71], executed the first TDD-SDR BS on a server and also, the first functional model of a Virtual Base Station (VBS) pool was presented. The outcomes satisfied the demands of the system used at that time [71]. Likewise in 2010, Bhanage et al., introduced a proposal for a virtualization wide-area network for 4G devices in a cellular BS to empower users participating among multiple autonomous slice users [72]. By 2011, the design of a mobile system that could reconfigure itself called Reconfigurable Mobile Network (RMN) has been put forward by Khan et al., they employed virtualization techniques to distribute the resources of networks [22]. In 2012, a design of Network Virtualization Substrate (NVS) for base station uplink and downlink sources split into parts was provided by Kokku et al., two situations have been adopted: firstly, they showed the optimal slice list with resources and bandwidth bases. Secondly, they customized the flow scheduling within the base station on a single-slice basis [43]. By 2013, the importance of isolation resource requirements being provided to BSs had been grasped by Kiess et al., focused on OFDM as a wireless interface, preparing for the virtualizing of an OFDM base station [73]. In 2014, a structure for virtualizing the Long Term Evolution (LTE) BS was introduced to improve the subscribers' performance and decrease the interference of the downlink beacon. They employed the assets of OpenFlow and SDN. The outcomes revealed improvements in the throughput, the average interference, and lost packets transferred between End-to-End users [74]. Nakauchi et al., suggested a virtual base station, with the capability to produce dynamic virtualized wireless communication networks. They utilised a couple of techniques, first, a VBS configured dynamically and controlling several physical wireless Access Points (APs) by applying the OpenFlow features. Second, seamless handover is handled between VBSs pre-authentication and pre-association at the target VBS [75]. Xiaodong et al., utilised FNA within SND for next-generation mobile networks with MOFP have been provided. Furthermore, they introduced an adoption strategy to concentrate on the User-Centric scenario in FNA [76]. Dawson et la., depending upon C-RAN, suggested a VBS network structure that enables connection to the CN [77]. In 2015, Wang et la., proposed virtual base stations, where every cell is dynamically determined by allotting virtualized signal sources to the fronthaul link for both the RRHs and BBUs. The functional process of the BBU is cloudified and virtualized according to the functional entities performance [78]. The considerable advancement of industries of electronic communication devices, especially in relation to fixed, wireless, and mobile devices production making these devices in tiny sizes. The Field-Programmable Gate Array (FPGA) technology is a hopeful advancement for solving the needed elasticity for producing and implementing 5G and beyond mobile communication networks. FPGA assists in the LTE scheme virtualization by carrying out a particular task in the CP. This usage can be observed in multi-mode base station infrastructure devices [79]. Virtualizing a single entity of a CN was introduced by Heinonen et al., who worked on the mobility and latency of 5G network. They brought forward the Mobility Management Entity (MME) to be set in a virtualized network located between the radio access network and the CN [80]. In [81], a logical architecture for 5G systems is proposed, in the form of a scheme for managing mobility among different access networks based on SDN and NFV. However, to date, the current communication networks, especially wireless and mobile networks have not wholly been virtualized [73, 80].

3.3 The Proposed System

Mobile communication systems suffer from End-to-End latency due to different types of delays such as RF registration delay, control delay, transmission delay, and processing delay. Minimizing these can be achieved by adopting new procedures that manipulate the establishment and negotiation of a contact session. The servers (BBU and CN entities) are responsible for making decisions of an End-to-End connection based on executing the function programs. These servers are usually placed away from the User Equipments (UEs). To decrease the End-to-End latency, we proposed a novel BS that contributes to making connection decisions through virtualizing the functions of the BBU and the CN servers and administering these functions near the UEs. The next section presents an explanation and illustration of the proposed system.

3.3.1 EPS Functions Slicing and Virtualising

The emergence of new technologies is helping in the design and creation of different types of network topologies. These technologies such as SDN, NFV, network virtualization, network slicing, network cloudification, and other virtualizing techniques, are facilitating the network service providers and operators in eliminating the subordination of traditional physical network infrastructure elements and the dedication of physical servers for each function[69]. The functions of these servers can be performed by adopting an appropriate virtualization technique for emulating in making decisions that are yielded by completing the execution of those functions. These *decisions* can be defined as the outputs of those functions that are implemented by the BBU pool and the CN entities, which follow software or programs that control the functions. Because the EPS system makes decisions (outputs of function) to establish, create and maintain a connection between End-to-End according to those functions. Therefore SVeNBs' VMs host the kernel of those functions to make its decisions regarding the local UEs. Examples of the EPS functions that are hosted by SVeNB are:

- Policy Control and Charging Rules Function (PCRF): detects the service flows, determines how data stream moves, and it is responsible for charging policy;
- Packet Data Network Gateway (P-GW): is responsible for assigning IP addresses to UEs, implementing a QoS application, and following PCRF per flow rules;

- Serving Gateway (S-GW): creates, deletes, and modifies the bearers for each UE that connected to the EPS. Performing of these functions depends on per Packet Data Network (PDN) established connections for each UE;
- Mobility Management Entity (MME): achieves the high-level signaling of the UE , the issues of authentication and security, besides to the management of data streams that unrelated to radio communications.

Functions slicing is defined as partitioning a function between two servers or VMs to complete a specific task to make decisions, one of which is complementary to the other's functionality. One of these servers represents a *MASTER entity, which is located at the CN*, whereas the second is a *SLAVE installed in SVeNB as a VM*. The SLAVE entity receives a function task pre-processed in the MASTER entity, that is, the former acts to complement the function already pre-processed by the latter. While *functions virtualizing* technique supports the possibility to virtualize any task or function that could be achieved by software programs, to be executed through a VM to produce session establishment decisions. i.e., functions of the BBU pool and CN servers such as MME, S-GW, P-GW, and PCRF as traditional off-the-shelf servers could be logically virtualized their functions by suitable VM's configuration [17].

3.3.2 Architecture and Functions of the *SVeNB*

Any virtualization of a network device needs three main physical hardware elements to build and implement a VM: CPU, RAM, and storage (HDD or SSD). Moreover, three main software elements, including an Operating System (OS), application programs, and data need to be considered. The SVeNB has been proposed to consist of one small device has CPU, RAM, storage, besides to peripheral slots allow to insert FPGA boards, which could be added for achieving at least one of the EPS server functions. Each slot represents at least a single VM depends on SVeNB physical hardware and shares its resources. The SVeNB has three main interfaces, the first being connect to the BBU pool and CN, whilst the second is to connected to the Internet gateway and the third interface connected to the RRH. A hypervisor is used to coordinate the VMs with physical resources. Figure 3.1 shows the proposed architecture of the SVeNB.

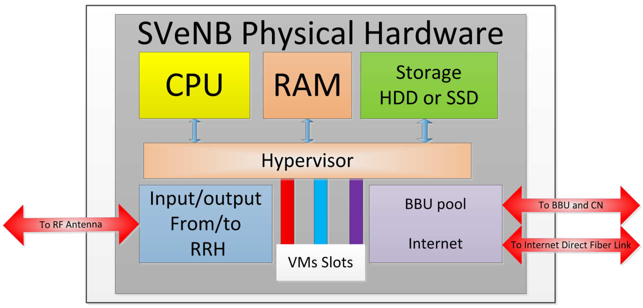


Fig. 3.1 The proposed architecture of SVeNB with peripherals VMs slots

The functions of the SVeNB are that it will have the ability to achieve a connection between the local subscribers or to the Internet directly. It can perform this connection without waiting for the decisions that are usually made in the EPC to create the requirements for the session. SVeNB can serve the subscriber who attempts to make this session or to connect to the Internet. This is based on the requirements that are pre-processed in the BBU pool and CN for all local subscribers covered by SVeNB. All user profiles are saved in the storage of SVeNB as cached user profiles. According to the user information profiles, SVeNB controls and governs the connection conditions between local subscribers. It takes part in making the decisions that are pre-negotiated with the BBU pool and the CN. It saves these decisions to be applied to make the final ones that enable the subscribers to communicate with each other. Also, it reserves pre-emption of bearer radio resources (ready to use), which are assigned by SVeNB to the users who want to connect with EPC without asking the BBU pool and CN again. The allocated IP addresses by P-GW are stored in SVeNB as ready to use. The SVeNB's VMs can detect and serve the type of required service, such as assigning an IP address. Consequently, based on this service detection the controller can decide to give QoS to the user. Moreover, it selects the interface to flow the classified data. Depending on the rules of PCRF (operates in real time in order to determine policy rules in the CN), the controlled flow-based charging functionalities in the Policy Control Enforcement Function (PCEF) that are virtualized and installed on the VM that works as the PCEF virtual server (supports offline and online charging interactions while PCRF does not support these). This virtual PCEF is responsible for dedicating an appropriated link connection, charging session establishment, and maintaining the established connection. The LTE and EPC functions that determine the decisions of a connection establishment are sent to SVeNB, which stores these decisions to be used for resolving the final decisions of the contact establishment.

This procedure will lighten the load on the BBU pool and the CN. Furthermore, this will decrease the End-to-End delay, and expedite the process of initiating the communication connections, in particular, when the local users attempt to link with others or to the Internet.

3.3.3 $UE - eNB$ Connection in C-RAN and *SVeNB*

In LTE, the Random Access Channel (RACH) consists of 64 Random Access Preambles (RAP) allocated for each cell of an eNB and there are two ways to get RACH by a UE: (i) Contention-based, where any UE can access this when in need of an uplink connection; and (ii) Contention-free RACH can be used in cases, such as handover and downlink data stream, where low latency is required [82]. The message exchange between the UE and eNB is either for location area updates or for initial access to set up a connection. In the latter case, the UE sends a request via a RAP message on the Physical Random Access Channel (PRACH) resources associated with Random Access Radio Network Temporary Identity (RA-RNTI) to the eNB, which covers the UE. The eNB replies with a Random Access Response (RAR) on the Physical Downlink Shared Channel (PDSCH), which is addressed by an ID and then, RA-RNTI is sent. The UE sends Layer 2 and Layer 3 messages, this information is the first scheduled uplink transmission on the PUSCH and formulates the use of a Hybrid Automatic Repeat Request (HARQ), including carrying of UE identifiers. This is the actual random access procedure message. The last action is the contention resolution message, which the UEs respond to on reception in two possible ways. The first is through a Positive Acknowledgement (ACK), which indicates that the UE has accurately decoded the message and thus, detected its own identity. The second probability is Discontinuous Transmission (DTX), which means that either the UE has correctly decoded the information and discovered that it contains another UE's identity (contention resolution), or it cannot decode the message. Figure 3.2 illustrates the random access procedure [83, 84].

At this point, the proposed SVeNB works as legacy eNB, based on the local user profiles that are saved by its storage unit. The installed VMs check the user profiles to see who needs service and then, it takes a decision (to begin the establishment of a session or not), according to having checked the user charge allowance, authentication, and other user profile matters. Subsequently, it assigns an IP address to any user who needs to make a call or contact the Internet. The connection between SVeNB

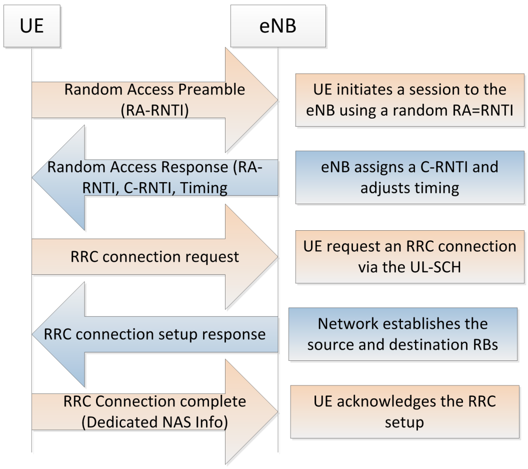

UE
eNB
Random Access Preamble (RA-RNTI)
UE initiates a session to the eNB using a random RA=RNTI
Random Access Response (RA-RNTI, C-RNTI, Timing
eNB assigns a C-RNTI and adjusts timing
RRC connection request
UE request an RRC connection via the UL-SCH
RRC connection setup response
Network establishes the source and destination RBs
RRC Connection complete (Dedicated NAS Info)
UE acknowledges the RRC setup

Fig. 3.2 Contention-based random access procedure

and the Internet can be made in two ways. The first, is via the direct fiber link that is connected directly to the Internet through the operator's S-GW, thereby bypassing the BBU pool and CN. This path is the default way to reaching all IP networks. The second way is through the BBU pool, CN, and Operator's S-GW, which is a backup path for connecting to all IP networks. The default path mitigates the load on the BBU pool and the CN, expedites the data delivery to the users, decreases end to end latency, increases the amount of delivered data by end users, and decreases the utilization of network bandwidth for control messages

3.4 End to End Delays Analysis

In this section, discussion on End-to-End delays or session establishment delays is presented. We compared the proposed SVeNB and C-RAN structure connected by the fiber in the fronthaul and backhaul links, concentrating on the End-to-End delay analysis. The scenario suggests that two local UEs try to connect to each other within one RRH. We compared our proposal with the C-RAN system to initiate connection steps. Figure 3.3 shows the representation of the C-RAN steps of session establishment verification. In step 1, the user contacts the RRH to get the radio bearer and IP address, whilst Step 2 the RRH asks the BBU for this request. The BBU processes the measurements of digital signaling and forwards the request to the CN to complete the session establishment. During Step 3, the CN performs the principal proceedings, such as determining the QoS for the bearer, IP distribution, and allowance charging, etc. This requires all the source CP messages crossing through the BBU and CN. Steps 4, 5, and 6 describe preparing to communicate with the target UE.

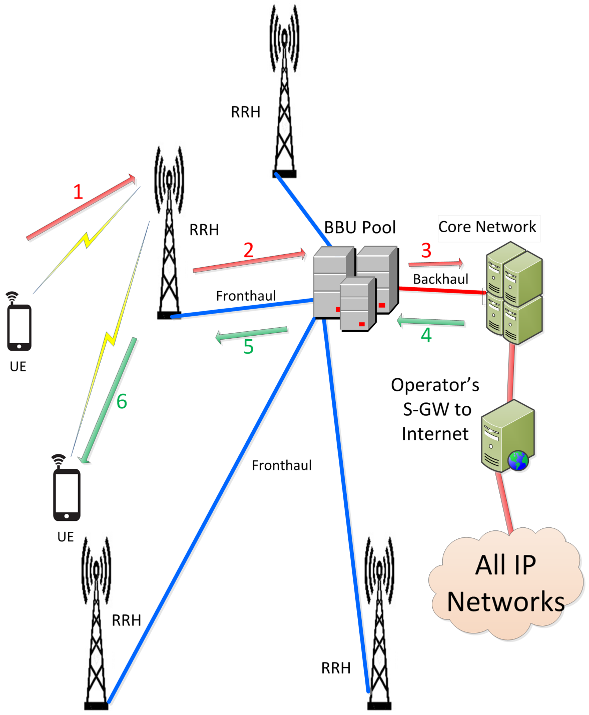


Fig. 3.3 The C-RAN End-to-End CP session establishment

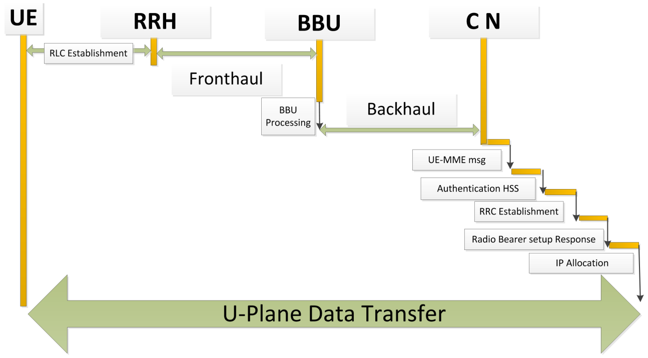


Fig. 3.4 The C-RAN End-to-End CP processing steps to establish a session

Figure 3.4 represents the principal steps of processing to set up an End-to-End association. The transmission delay between the RRH and the BBU (fronthaul link) as well as the BBU pool and the CN (backhaul link) relies upon the separation distance of the connection between them. In the CN, there are several servers or entities executing the jobs of establishing the End-to-End session, with each having its processing and queuing delay times. The connection steps above mentioned become more complicated if the users are under different RRH coverage, for, the BBU and the CN make a check on each user to initiate the connection between them. Consequently, in such a case the End-to-End delay will increase due to need for more processing to achieve the session connection. However, with our proposal, the session connection is manipulated and treated locally in SVeNB, which initiates the connection based on the information received from the BBU pool and CN.

SVeNB accomplishes the tie without attaching to the BBU pool or the CN again. The idea is a creative step because of the semi-centralized nature of initial connection processing and the capability of SVeNB to make decisions for the UEs as a semi-standalone base station. Figure 3.5 explains the structure of this scenario. The procedure of this scenario as follows.

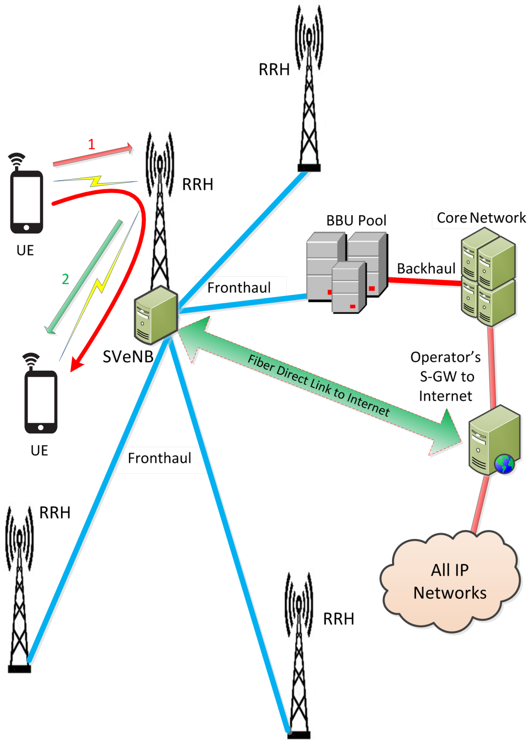


Fig. 3.5 The Proposed SVeNB End-to-End CP session establishment

- Step 1 depicts the UE's request to obtain a bearer and IP address. SVeNB responds to this request by checking the user profiles stored in its storage unit. Then, the installed VMs of SVeNB can allot and assign the bearer and the IP to that UE. These VMs perform the CP measurements of this session, such as authentication, security, charge allowance, and IP assignment.

- Step 2 depicts the association with the destination UE to begin the session and exchange data between it and the source.

Figure 3.6 explains the procedure for achieving this association. When the user contacts SVeNB, the latter checks the location of the UE by the virtual MME (VMME) to verify the position and authentication through the interaction with the virtual HSS (VHSS). A virtual machine (P/S-GW) acts as a gateway or mobile IP anchor of SVeNB to the Internet, with the radio bearer and IP address being assigned in this stage. The connection of two users that are covered by a different coverage area of RRHs is easily prepared and executed by SVeNB. This is owing to the information of RRHs, such as frequency bands and IP addresses, having been saved and ready to use by SVeNB. Hence, SVeNB can easily overcome the complication of communication between two supported by different RRHs. In other words, users do not need to contact the BBU pool and the CN to allow for session establishment. As a result, the End-to-End delay time decreases due to no more demand for processing to achieve the session connection.

At the end of these steps, the user has two choices. The first, is to connect to another user, which is also covered by the same SVeNB or it covered by another one. The second choice is to connect to the Internet via the direct fiber link, which has been proposed to connect SVeNB directly to the mobile operator's S-GW in case the UE requires surfing of the Internet, besides to the regular path (SVeNB-BBU pool-CN-S-GW path).

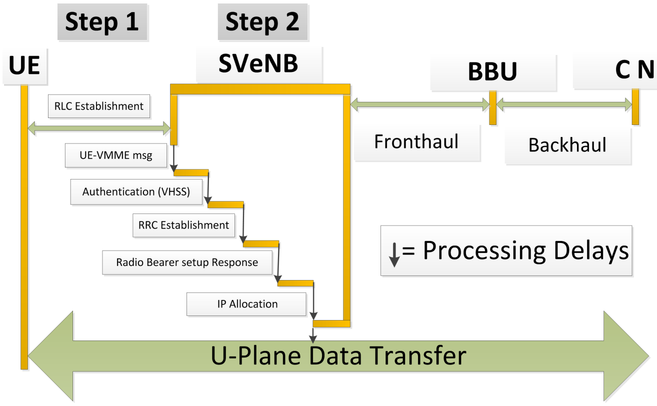


Fig. 3.6 The Proposed SVeNB End-to-End CP processing steps

The assumption in the C-RAN system, the delays time of a transmission between a UE and RRH is t_{LRC} which implies the radio link control delay. The link delay between the RRH and BBU pool which stands for the fronthaul delay is (t_F), and the delay between the BBU pool and the CN as the backhaul delay is (t_B). D_T is the summation of the transmission delays of the radio bearer of the RRH, the fronthaul, and the backhaul links. Hence, the total transmission delay can be represented by the Equation 3.1,

$$D_T = t_{LRC} + t_F + t_B \tag{3.1}$$

The processing delay time in the BBU pool is t_{PBB}, which expresses the overall time that is consumed by the BBU pool for the CP signal to initiate the association. The processing delay time that is spent by the CN is t_{PCN}, which denotes the overall time that is consumed by the CN for control information data to establish that connection. Hence, the summation of the processing delays of the BBU pool and CN is D_P. It can be estimated by,

$$D_P = t_{PBB} + t_{PCN} \tag{3.2}$$

D_Q is the summation of queuing delay time in the BBU is t_{QBB}, and the queuing delay time in the CN is t_{QCN}. Consequently, D_Q can be expressed by,

$$D_Q = t_{QBB} + t_{QCN} \tag{3.3}$$

Consequently, the total delay time D can be expressed with the following equations,

$$D = D_T + D_P + D_Q \tag{3.4}$$

The processing and queuing delay times in the RRHs are very small, so it can be ignored. All these time delays are introduced by the C-RAN and the CN just for a session establishment of End-to-End association (i.e. only CP messages exchange). Whereas with SVeNB, the equivalent connection can be made, but with the elimination of D_T. This is because this delay time has already been consumed by the first attachment of the BBU pool and the CN by SVeNB when the profiles of the subscribers are manipulated and sent to the SVeNB. The processing and Queuing delay times are reduced because they are processed by one physical device through many VMs supported by SVeNB. This leads to a decrease in the period of the delays time of the processing and the queuing. Therefore, to compare between the SVeNB and C-RAN, it should be taken into consideration those facts. With SVeNB the link delay time will be dropped from the delay calculation, whilst the processing and queuing delays should be estimated as a VM. Hence, the delay time equation for SVeNB will be,

$$D_S = D_{PS} + D_{QS} \tag{3.5}$$

where, D_S is the total delay time in SVeNB, D_{PS} is the processing delay, and D_{QS} is the queuing delay in the SVeNB. The D_{PS} and D_{QS} notably rely on the specifications of the SVeNB physical hardware and capital expenditure can decrease this delay through hardware with high specifications.

3.4.1 Radio Link Delay Analysis

According to 3GPP TS 36.213 and TS 36.306, for transmission in mobile networks, the Radio Link Control (RLC) has been using in the LTE to enhancement the Frame Error Rate (FER), according to the condition and availability of the Bandwidth (BW) channels. However, we assume using different BW for LTE (5, 10, 15, and 20 MHz) and then, the capacity of each channel can be found by a simple calculation. Consequently, the number of frames per packet (k) can be calculated for each channel rate. The LTE frame duration τ equals to 10 ms. Also, assume modulation scheme QPSK is used. The 1526 bytes is a Maximum Transmission Unit (MTU) for IPv4 packets. While IPv6 can convey much larger packets that make it well matched with the LTE system, which can handle the large size of packet or frame. The LTE frame is known as a Transport Block (TB) and it contains numerous sub-frames. Every sub-frame consists of two slots, each of which has six or seven OFDM symbols according to the extended or normal cyclic prefix, respectively. The number of bits in each symbol depends on the modulation scheme. In LTE, the maximum control message size does not exceed 10% of the sub-frame that is assigned to the UE [85], with the number of bits (b) in one sub-frame being:

$$\begin{aligned} b =& 2(RB) \times 12 \textit{ (sub-carriers)} \\ & \times 7 \textit{ (assuming 7 OFDM symbols)} \\ & \times 2 \textit{ (slots per sub-frame)} \times M \\ =& 336 \times M \end{aligned}$$

M represents the modulation scheme. If the QPSK is used for control messages, then M equals 2. So, 336 x 2 x 0.1 (10% as assumed above) $\approx$ 68 bits. Consequently, the number of bytes per Transmission Time Interval (TTI) can be calculated for the specific channel as the channel BW multiplied by TTI divided by 8, for instance, 10 Mbps x 10 ms x$\frac{1}{8}$ = 1250 Bytes. *This leads to the smallest transmission channel capacity that will cover the largest transmitted control message.*

Assume the probability for QPSK to transmit an RLC frame successfully is:

$$P_s = 1 - P(B_n) \tag{3.6}$$

The frames that are dropped or lost is denoted as B_n, and n is the number of re-transmission trial (typically 3). The probability of effective packet loss is:

$$P_f = P(B_n) = P[P(2-P)]^{\frac{(n+n^2)}{2}} \tag{3.7}$$

Assuming that the LTE channel operates under normal conditions, the probability of the FER is p = 10^{-2}, and n $\leq$ 3, the P_f can be calculated as follows:

$P_f = 10^{-2}[(2 \times 10^{-2} - 10^{-4})]^6 = 6.21 \times 10^{-13}$

From [86, 87], and [88], the typical value of the propagation delay is D_{pr} = 100 ms. Thus, the equation of RLC transmission delay can be calculated as,

$$\begin{aligned} t_{RLC} = & D_{pr} + \frac{[P_f - (1-P)]}{P_f^2} \\ & \times | \sum_{j=1}^{n} \sum_{i=1}^{j} P(C_{ij}) \\ & \times [2jD_{pr} + (\frac{j(j+1)}{2}) + i) \times \tau]| \end{aligned} \tag{3.8}$$

The first frame is received by end point properly is denoted as $C_(ij)$, at re-transmission ith of the transmitted frame at jth re-transmission trial.

3.4.2 Fibre Link Delay Analysis

The speed of light v decreases when it transfers into fiber optics cables due to the refractive index n_r of the material made these cables. The speed of light c in free space is 299792.458 km/s. In the single-mode optical fiber transmission, the standard wavelengths are 1310 nm and 1550 nm with refractive indices 1.4676 and 1.4682, respectively. Then, the delay time for a 1 km length of the fiber and a different refractive index according to a particular wavelength can be calculated by Equation 3.9,

$$t_\lambda = \frac{1km}{v_\lambda} = \frac{(1km \times n_\lambda)}{c} \tag{3.9}$$

The delay times for wavelength 1310 nm and 1550 nm are 4.895 μs and 4.897 μs, respectively [89]. So, for the Round Trip Time (RTT) for these values should be doubled. In the C-RAN architecture, the distance between the BBU and RRH is restricted by processing and transmitting delay times which should be not more than 3 ms (the period from uplink to downlink) [90]. In other words, when increasing the distance between the BBU and RRH, the processing delay of the former should decrease. The maximum distance of fiber optics cable that links the BBU pool and the RRHs can be given by solving the following equation

$$M_d = \frac{RTT \times c}{2 \times n_\lambda} \tag{3.10}$$

Equation 10 can be used to determine the maximum distance between the BBU and RRH with different kinds of optical fiber link for a specific λ.

3.4.3 Queuing Delay Analysis

Queuing delays between the sender and the receiver UEs depend upon the amount of data packets in every specific queue at each entity in the network. The queuing delay time for establishing the EPS session can be represented as the summation of the delays of all the entities involved in initiating the session of the CN network and the BBU queuing delays. AS these entities successively process its functions due to some of the entities (*Master*) administrator and control the other entities (*Slave*). In other words, some decisions determine the beginning or lunching of other function to make decisions. For example, the decisions regarding authentication and allowance charging functions are first made, then the other functions are implemented to make the final decisions to establish a session connection. The final decisions are yielded from execution the functions of diverse entities to establish this connection. As a result, it can be considered the M/M/1 queue assumption of CP messages at these entities. For each entity of the EPS system, the queuing model of the M/M/1 queue and Poisson signaling arrival rate process have been utilized by the proposal. For M/M/1, if the first input rule is Poisson, then the subsequent stage input is additionally Poisson and independent of the input process and so on [91, 92].

Assume Z is the number of entities (the BBU and the CN) included in session establishment, δ is the Poisson arrival process rate (packet/sec), and μ is the transmission arrival packet rate of the queue (packets/sec). Every element allows a traffic load of ρ_i = δ_i /μ, i = 1, 2, ... , Z. Moreover, in mobile communication networks some of the servers depend on others to make their decisions, i.e. the *ith* entity receives traffic from the leader (*Master*) entity (any chief entity could be either a leader or follower (*Slave*) as per the instance of the necessity of information preparing among them) with a packet arrival rate δ_i. Consider the queuing delay time at the recipient buffer only (for simplicity) [93]. The expected sum of the queuing delay time is the sum of the expected queues in tandem at every element, which can be represented as follows:

$$E[X] = \sum_{i=1}^{Z} E_i \left[\frac{1}{\mu_i} \right] \tag{3.11}$$

where, X = 1/μ is the service delay time of the entity, and μ is the transmission packet rate of the queue. These amounts of delays can be determined by formulas that are derived by the Markov chain.

Then, the possibility of there being packets in the queue is, $P = \rho^N(1-\rho)$, where, N is the number of packets that are sent by a channel, and $\rho = \delta/\mu$, from this we can get $N = \rho/(1-\rho)$. So, the delay time for one entity is,

$$t_Q = \frac{N}{\delta} = \frac{\rho}{\delta(1-\rho)} = \frac{1}{(\mu-\delta)} \tag{3.12}$$

For the C-RAN system, the total queuing delay time can be given as:

$$D_Q = \sum_{i=1}^{Z} \frac{1}{(\mu_i - \delta_i)} \tag{3.13}$$

Whereas in SVeNB the queueing model is M/M/m [93] because its VMs perform m jobs and consequently, the queuing delay time will be expressed as [63]:

$$D_{QS} = \sum_{i=1}^{Z} \frac{1}{(m\mu_i - \delta_i)} \tag{3.14}$$

3.4.4 Processing Delay Time

To estimate the processing delay time that is required to build up a session in the EPS, the time spent by each entity to complete its job should be determined. Authors in [90] mentioned that the BBU processing for one round trip is 3 ms. The EPC entities (servers) need to perform further processing on arrived packets, i.e. the entity should decapsulate the arrived packets to take a decision and then encapsulate them again. The processing delay time depends upon the packet length. Let L denotes the IP length (32 bit or 128 bit), M_s is the machine word size, W represents arrived word size, R is the user profile record size, and S represents the size of the server's processor architecture, e.g., 32 or 64 bits of the processor of an entity.

the time spent by each entity to complete its job should be determined by around 100 ns to finish this process. So, the processing delay time for one entity can be obtained from [94]:

$$t_P = 100\frac{W}{M_s} \times \left[log_{sys}R + \frac{L}{S}\right] \tag{3.15}$$

Next, we have derived the total processing delay time equation as:

$$D_P = \sum_{i=1}^{Z} \sum_{j=1}^{F} 100 \frac{W_{ij}}{M_{si}} \times \left[log_{sys} R_{ij} + \frac{L_j}{S_i} \right] \tag{3.16}$$

where, Z represents the number of entities of the EPS, F denotes the number of the user's messages to a specific entity. jth is user's parameters to be treated by ith entity, and sys is the system constant [63].

3.5 Performance Evaluation of *SVeNB*

This section discusses the results of the local UEs' End-to-End delay control messages, with consideration of both SVeNB and C-RAN. Figure 3.7 explains the relationship between the service rate and the End-to-End delay of the control signaling to establish a session for local UEs that are covered by a base station. At the service rate of 310 packets/second SVeNB performs with a delay time of almost 4.5 ms. Whilst in C-RAN it is slightly more than 9 ms. Both systems produce lower values in the delay time when increasing the average service rate. However, the SVeNB processing delay time becomes one-third of that of C-RAN at the service rate 400 packet/second. The performance of SVeNB for control plane signaling to establish End-to-End session is at least double of the C-RAN in the case of the low value of average service rates. The delay time reduces with increasing the serviced packets rate. When initiating the session, SVeNB and C-RAN almost work in a similar way (i.e. the initial access to set up a session establishment), with each executing almost the same procedure to set up the RACH for every user. However, following the setting up of the RACH the delay time for SVeNB becomes much less than with C-RAN due to the former not needing to connect the BBU and the CN, because all the information that it requires to achieve a session is already stored into its storage unit [63].

Fig. 3.8 explicates SVeNB's superiority over C-RAN when serving the same number of served user profiles. The figure exhibits that when a user attempts to establish a connection, both systems take a long time of processing in the beginning, due to having to prepare the associated data of the user to switch from the inoperative to operative state. Subsequent to that, both systems arrive in a steady state case and the processing delay time experienced with SVeNB is around 0.2 μs at 100 bytes of user profile size. In contrast, C-RAN consumes approximately 0.6 μs for the same user profile size.

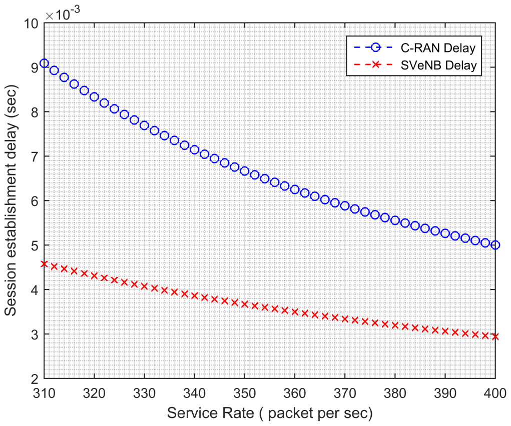


Fig. 3.7 Session establishment CP delay with service rate comparison

Expanding the size of the user profile in the range of 100 times will provide an additional delay time of about 0.1 μs in both systems for the steady-state case.

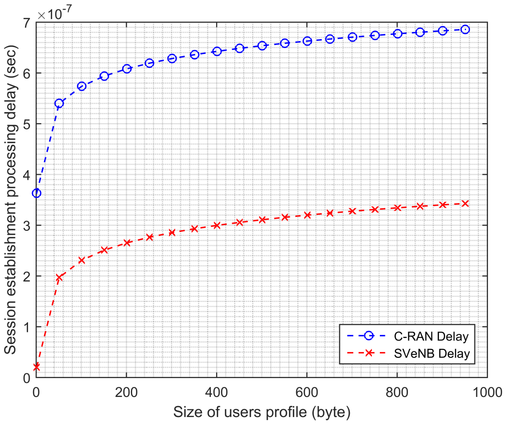


Fig. 3.8 Session establishment processing delay and user profiles size

In C-RAN, the BBU and the CN serve several cells and hence, the processing delay time grows according to the increase in the number of served cells. As a result, the processing delay time in C-RAN is greater than SVeNB, this being due to the massive number of user profiles and growth in their size. Intuitively, the base station should wait for more time to be served by the C-RAN structure.

Fig. 3.9 shows the average delay versus packet arrival rate for a single user. Considering the transmission, queuing, and processing delays, SVeNB provides higher performance than C-RAN in delivering data to a user for the same circumstances. This means that the data rate is increased as the total delay (transmission, queuing, and processing) is decreased, when comparing the attainment

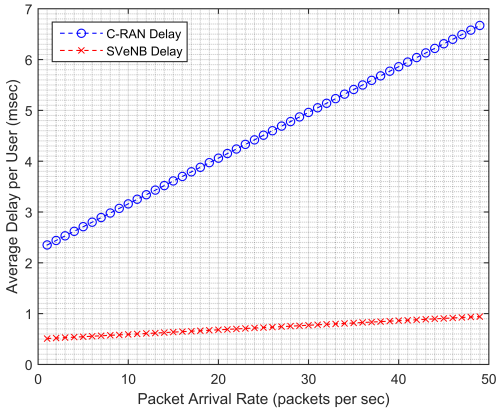


Fig. 3.9 Average Delay per User against Packets Arrival Rate

between SVeNB and C-RAN. Moreover, the average delay in C-RAN expands dramatically due to the accumulation of different devices which cause delay, while SVeNB does not suffer from such delays. This fact is true if we exclude the first contact to the BBU and the CN by SVeNB for acquiring the profiles of the users. The provided service rate of a network is influenced by the number of users

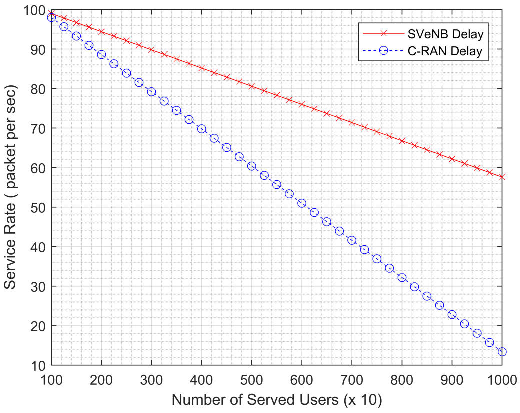


Fig. 3.10 Impact of Increasing number of Users on Service rate

that are served by it. That is, the service rate decreases with an increasing number of users in the network. This degradation is the result of the numerous operations performed by separated physical machines in the C-RAN network to perform a particular task. SVeNB overcomes this obstacle through virtualizing and mimicking these separated servers and functions, whilst gathering them into a single physical device. Fig. 3.10 presents the service rate versus the number of users. Also, the figure shows the performance of the proposed SVeNB against C-RAN network. In practical life, UE needs

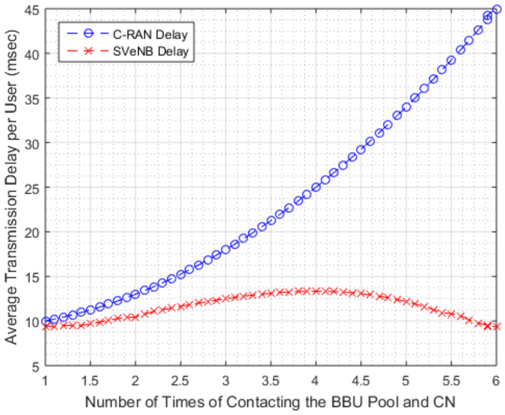


Fig. 3.11 Average Transmission Delay per User and Communication With BBU and CN

to communicate with the BBU and the CN to perform the important measurements to initial the session. Fig. 3.11 illustrates the average delay of transmission links that connect the RRH, BBU, and CN. At the starting point of the graph, both SVeNB and C-RAN required the same time to begin a session. SVeNB has the ability to handle and process the requirements of a session without the need to communicate with the BBU pool and CN again. Conversely, the C-RAN network cannot initiate a session without contacting them more than once and thus, the line graph of SVeNB drops to a minimum value, whereas that of C-RAN rises to a maximum one.

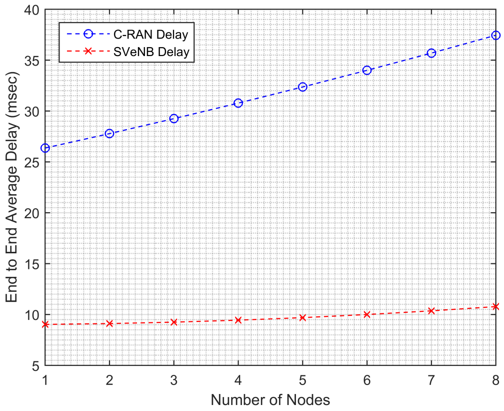


Fig. 3.12 Average End to End Delay

Fig. 3.12 exhibits the End-to-End average delay comparison between SVeNB and C-RAN. This increases as the number of nodes increases due to each node (physical or virtual entity) suffering from its own delays, such as transmission, queuing, and processing. SVeNB can handle multiple criteria for user profiles at the same time and on the same server. This procedure leads to a reduction in the End-to-End delay time required to establish a session between users. Moreover, the End-to-End average delay in C-RAN grows dramatically due to the successive delays of different devices that include establishing a connection. While SVeNB does not suffer from the accumulative delays as well as in C-RAN which needs to acquire the profiles of the users from different servers.

3.6 Chapter Summary

This chapter aimed to present a novel base station idea based on virtualization technologies and the End-to-End delay time has been analyzed. The outcomes indicate that the delay time to establish the End-to-End association by the proposed SVeNB is less than that by the C-RAN. Mobile networks need to make substantial progress to meet the requirements of 5G networks and beyond. This requires more interaction between the new technology of the physical infrastructure and VMs to achieve complete virtualized networks. The most significant delay time is consumed by the processing of the arrived control data packets from/to the UE and CN for establishing the End-to-End link. Most notably, the processing delay of the control plane messages is where the highest delay time occurs. Moreover, control plane signaling in best conditions requires more than 60 ms to establish a session. The proposal concentrates on lowering this delay time by virtualizing and slicing the primary functionalities of the BBU pool and CN. Moreover, bringing these functions close to the end user enables SVeNB to produce a significant reduction in control plane delay time. A comparison has been performed with the results that were obtained by SVeNB and the C-RAN emulating approach. The comparison involved variant values of average service rates. SVeNB exhibited around 62% less delay time than C-RAN. In addition, the performance of SVeNB is higher than of the C-RAN by 68% in terms of user profile processing. The recommendation is that a considerable solution to overcoming the End-to-End delay problems is by virtualizing and slicing functions of the BBU and CN. Also, this requires bringing those virtualized functions close to the end user. The semi-decentralization of the processing functions mitigates the load on the BBU pool and CN.

Chapter 4

Smart Virtualization for Packet Forwarding

4.1 Introduction

The most important feature of modern life is to be connected to the Internet whether by fixed or mobile devices. Smartphones are the most significant of these devices which burst into a broad spread in whole the world. As a result, the IPv4 addresses are depleted due to the growing number of the smart devices. Consequently, Internet service providers can not provide enough IP addresses to the continually increasing the number of that devices. By using IPv6 which can accommodate a tremendous amount of IP addresses that can be provided to all devices and networks that associated with the IP address [95]. The IP addresses enable the communication between mobile devices and their wireless access points. Therefore, wireless devices can identify themselves through their IP addresses. Moreover, the IP address is used as an indicator of the mobile device location as well as a device identifier. Also, it binds the mobile device and applications to their current location in Internet networks. The enormous increase in the number of mobile devices makes the binding is difficult to support the mobility through the Internet and the wireless networks due to increasing the size of the routing table, depletion IP addresses and bandwidth, more power consumption, and packets handover delay.

There are several protocols for mobility based on separation of the mobile IP into Locator and Identifier. In [96] explained the Mobile IPv6 (MIPv6) mechanism. This protocol used a permanent IP address called Home of Address (HoA) for an MN as an identifier and changeable Care of Address (CoA) as a locator. These routable IPs are managed by Home Agent (HA) which maintains the mapping between the HoA and CoA. Another protocol called proxy MIPv6(PMIPv6) [97] uses the Home Address (HoA) for identifying the MN and proxy CoA (PCoA) for the locator. Managing the binding information of HoA and PCoA is done by local mobility anchor (LMA) [98].

In last decade several protocols emerged based on separation of IP address into locator and identifier. These protocols suggested two concepts of separation. The first one is host-centric such as Host Identifier Protocol (HIP) [99], Site Multihoming by IPv6 Intermediation (Shim6) protocol [100], Mobile-Oriented Future Internet (MOFI) [101] and Locator-Identifier Separation Protocol host (LISP-host) [102]. The second concept is network-centric such as Locator-Identifier Separation Protocol Distributed Mobility Control (LISP-DMC) [103], and Distributed Hash Table (DHT)-based identifier-to-locator mapping [104]. All these protocols use tunneling techniques to deliver packets. These tunneling techniques deplete a significant amount of networks bandwidth.

The future communication networks should support the 5G mobile networks which promise to meet the 5G mobile network requirements such as i) low latency (less than 1ms); ii) high-speed mobility up to 500 km/h; iii) traffic density up to 1000 folds than today; iv) almost 100% coverage; v) less network management and administration; vi) separation of the DP from the CP of the network traffic; vii) flexible sharing of network resources. At the same, these networks should not increase the infrastructure cost and the power consumption. Besides, the self-organization network functions should present to manage the networks [105]. The thesis focused and worked on the delay reduction issue of handling the packets pass through the network devices during the handover for 5G networks.

To address the problems mentioned above, the proposed a mobile network adopted network-centric based on SDN and NFV. The virtualization technology actively empowers the SDN. The fundamental architecture of SDN network depends on decoupling the CP (which represents controlling packets that are created by the SDN controller) from the DP (which involves physical switches with high performance to deliver a pure data packets of a network). The CP can be implemented by a physical or virtual machine away from the DP [106]. Moreover, in the SDN environment, the control packets do not utilize the standard IP routing only, because it could use different algorithms and mechanisms according to which task is wanted to be implemented by the algorithm. The idea of this work is based on our previous proposal that entitled Smart Virtual eNB (SVeNB) [63]. The SVeNB has suggested of using SDN and NFV. The SDN harmonize with NFV technology which enables building a new virtualized mobile networks underpin by employing virtualization technology standards to consolidate the different network devices.

1. Creating a new identifier for an MN as a local identity within the mobile operator networks.
2. Adopting continuous, seamless packets delivery throughout the handover and a new mobility management mechanism based-SDN and NFV.

4.2 Background and Related Works

Reducing the handover and mobility management delays in mobile networks has been researched with different proposals. The most of these proposals suggested modification either in hardware such as access points or software like protocols [107]. To review the related works, the related works focused on the papers and articles that were proposed the separation of IPv6 address into locator and identifier concept and binding this concept with SDN to present our idea. Shim6 is a host-based multihoming layer 3 protocol. It provided more than one IPv6 addresses for a host, i.e., locator agility of fall over the ability for IPv6 nodes. A host employs shim6 can use more than one prefix of IPv6 addresses if the host has more than one interface for networks attachment [108]. HIP proposed a new layer called host identity layer. Host identity layer was inserted between layer 3 and layer 4 to identify the host. This layer maintains the mapping information of identifier and locator [99].

MOFI proposed local locator (local LOC) and HID for recognizing the locator and host. It proposed ID-LOC mapping to maintain HID-LOC information. Binding between the HIDs and LOCs was achieved by access router which has a hash table, an HID-LOC register, and local mapping controller [109]. LISP-MN supports the node mobility. Two subsets of a standard are used called Egress Tunnel Router (ETR) and Ingress Tunnel Router (ITR) functionality in an MN. A centralized mapping in LISP-MN was performed by a server who works as a mobility anchor for providing information on ID-LOC mapping. Shortcomings in LISP-MN are a double encapsulation required and triangle routing which caused the problems of path stretch [102]. Authors in [110], proposed an improvement to solve the problem of double encapsulation by localized the local LOC and Local Map Server (LMS) of mobility controller. All the protocols aforementioned are host-based. These protocols needed to modify the MN as hardware or software. Besides, they were difficult to deploy on the mobile networks. Moreover, they provided a single point of failure due to utilizing the centralized map server [111].

The second group of protocols adopted the network-based method. LISP has used a mechanism for alternating the IP addresses with two separated namespaces. First part is the Routing Locator (RLOC) which was used for Internet networks infrastructure. The second part was the Endpoint ID (EID) which was used by sites. LISP has mapped EIDs to RLOCs through the Map-Resolvers and Map-Servers [112]. Scalability and flexibility of packets routing can be provided by LISP-AR-DMC. Also, it solved the LISP-MN issues. The Access Routers (ARs) had a Tunnel Router (TR) function. Multicast communication was used among the ARs for mapping the ID-LOC. The DHT based on a resolver LOC/ID mapping approach has been suggested to solve the problem of the locator of a flat ID. DHT supposed every autonomous system manages the EID-LOC mapping information. It utilized a modified Content Addressable Network (CAN) which was applied "keys" onto "values" mapping [113]. Items are registered by the resolver to a CAN to refer that the EID-LOC mapping. However, all these protocols based host or network used encapsulation and tunnelling which caused depleting bandwidth, increasing power consumption and demanding more processing due to overhead data.

In mobile communications, the location of an MN must determine geographically and topologically to maintain the connection continuity. The mobile IPv6 address can be split into two parts. The first part is called prefix ID which represents the NID or topology ID (also known as locator ID) which associates with both the network and the subnet. Prefix ID consists of 64 bits. The least significant bits (16 bits) of that 64 bits was assigned to subnets and known as SID. The second part is the host location identifier which related to the IID or host identifier.

In each MN the MAC address is bound to each a specific interface of the network attachment point. So, all packets have to involve a MAC addresses of the source and the destination. The packets never across the link layer unless the MAC address is checked first. If the MAC address matches the packet is forwarded to the upper layers to verify that packet has delivered to the proper MN. The

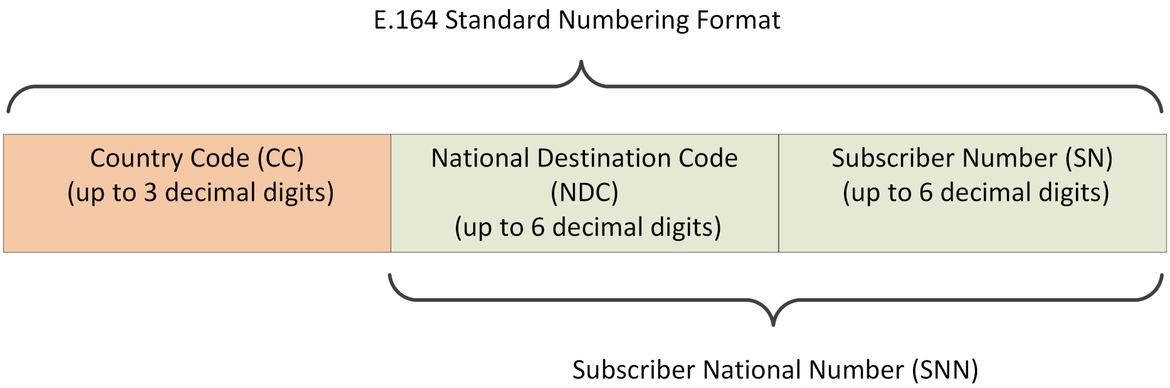


Fig. 4.1 The Structure of the E.164 Standard [114]

packet is dropped if there is no matching. This concept is the base of our proposed idea. The smart virtual eNB (SVeNB) is suitable to be the most candidate for our conception. SVeNB consists of several VMs. One of these VMs serves as S/P-GW, which serves the users within the coverage area of SVeNB.

The E.164 standard is defined by the International Telecommunications Union for Telecommunications (ITU-T). The ITU-T defines the international public telecommunication numbering plans and telephone number formats. The E.164 standard numbering can have a maximum of 15 decimal digits. The first part (one to three digits) of the telephone number is the Country Code (CC). The second part (up to six digits) is the National Destination Code (NDC). The last part (six digits) is the Subscriber Number (SN). The SN and NDC together are called the Subscriber National Number (SNN) [114]. Fig. 4.1 show the structure of E.164 standard.

4.3 Generating Mobile Node Tag

The proposal suggests a novel *identity* for an MN consists of the Organizationally Unique Identifier (OUI) which is 24 bits and the SNN. The SNNs can cover all probabilities of the mobile subscriber numbers. The likelihood of the most significant mobile number that can be formed from SNN (12 decimal digits) when all the digits are 9s, and that number can be represented by 40 bits. By combining the SNN and OUI, we can create a local identifier within the mobile operator networks. This identifier

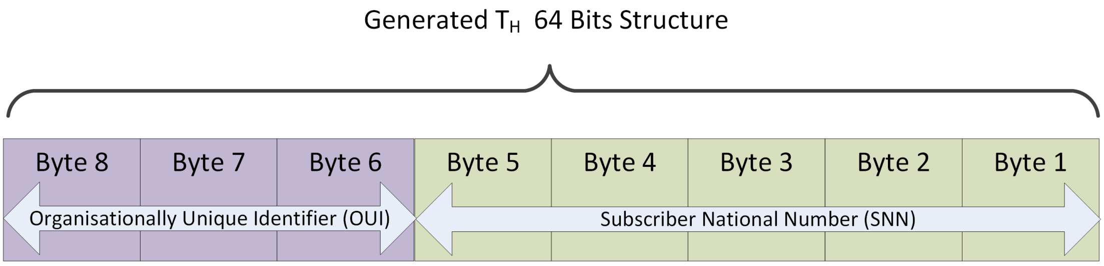

Generated T_H 64 Bits Structure
Byte 8
Byte 7
Byte 6
Byte 5
Byte 4
Byte 3
Byte 2
Byte 1
Organisationally Unique Identifier (OUI)
Subscriber National Number (SNN)

Fig. 4.2 The Structure Format of the Generated Tag

is 64 bits long, and it is compliant with IPv6. Moreover, it can be used as an alternative to IID part. Fig. 4.2 shows the format of generated tag.

Using the SNN and OUI as an MN ID is for security reason, also to distinguish the MN within specific mobile operator networks. The created tag is the objective to achieve local identifier (T_H) for operator networks only. i.e., T_H is well known for mobile operator network devices (SDN controllers, OpenFlow switches, and SVeNBs). The generating T_H should be created by the MN itself. The proposed approach to creating the T_H is accomplished by combining the OUI part of the MAC address of the MN and the SNN. The generated T_H is used by the network as the MN's ID. T_H inherits the global uniqueness feature from OUI and local uniqueness from SNN.

The process of generating T_H is done by the MN once only. Each MN retains its generated T_H. When the MN moves amongst the SVeNBs, it uses its T_H to contact with the new SVeNB as MN identity. T_H remains the same for that MN as long as it belongs to the same mobile operator networks. T_H must be changed if the SNN is changed. In this case, a new T_H should be generated by the MN with the new SNN.

The generated tag is unique due to the uniqueness of both the OUI of the MN and the SNN in the domains of a specific mobile operator network. In other words, the T_H is permanent for an MN even when it moves amongst the mobile operator network domains. T_H is not used by any router located outside the domains of the mobile operator networks. Therefore, packets are routed by the standard IPv6 protocols on the Internet networks and other networks which located outside the mobile operator domains.

4.4 Proposed System Architecture

The proposed system architecture as shown in Fig. 4.3 consists of two main parts. The first part represents the domains control layer. The second part is the domains themselves. Domains control layer comprises an Edge SDN Controller (ESDNc) which controls the Edge OpenFlow Switch (EOFS) and other Local SDNc (LSDNc)s and OpenFlow switches (OFS)s in each domain. Based on the information saved in ESDNc lookup table about the links which connect all devices involved in the system. The ESDNc governs the traffic flow from/to the mobile operator network through the EOFS and OFSs. According to this information ESDNc constructs and maintains its lookup table and makes the rules and actions which are sent to EOFS and other LSDNc and OFS.

The principal tasks of the ESDNc are the mobility management of MNs amongst the domains. Also, ESDNc directs the flows of data traffic which enters the EOFS (notably, in case of packets handover). Besides to the ESDNc, this layer comprises an EOFS which acts as edge point of aggregation and distribution of data streams from/to the OFS in each domain. The responsibility of the EOFS is to forward the packets based on the rules and actions in its flow tables. These flow tables are built and modified by the ESDNc.

The suggested network architecture for each domain consists of LSDNc, SVeNB, and at least one OFS. These domains are complementary with the most candidate 5G networks C-RAN architecture. The mobile network resources are virtualized to be hosted at SVeNBs. In the beginning, each SVeNB declares its link address to the LSDNc to receive packets of their users. The following sections explain the idea behind this proposed architecture.

4.4.1 Domains Control Layer

This layer contains ESDNc and EOFS devices. The ESDNc is connected to mobile operator core network, EOFS, OFSs, and LSDNcs in each domain. The principal duties of ESDNc are packets directing during the handover and mobility management of MNs amongst the domains and filtering the traffic. The ESDNc receives the information from MME, S-GW, and P-GW of the core network, LSDNcs, EOFS, and OFSs to update its lookup tables. According to that information, the ESDNc makes decisions (rules and actions) and sends these decisions to EOFS, OFSs, and LSDNcs. The

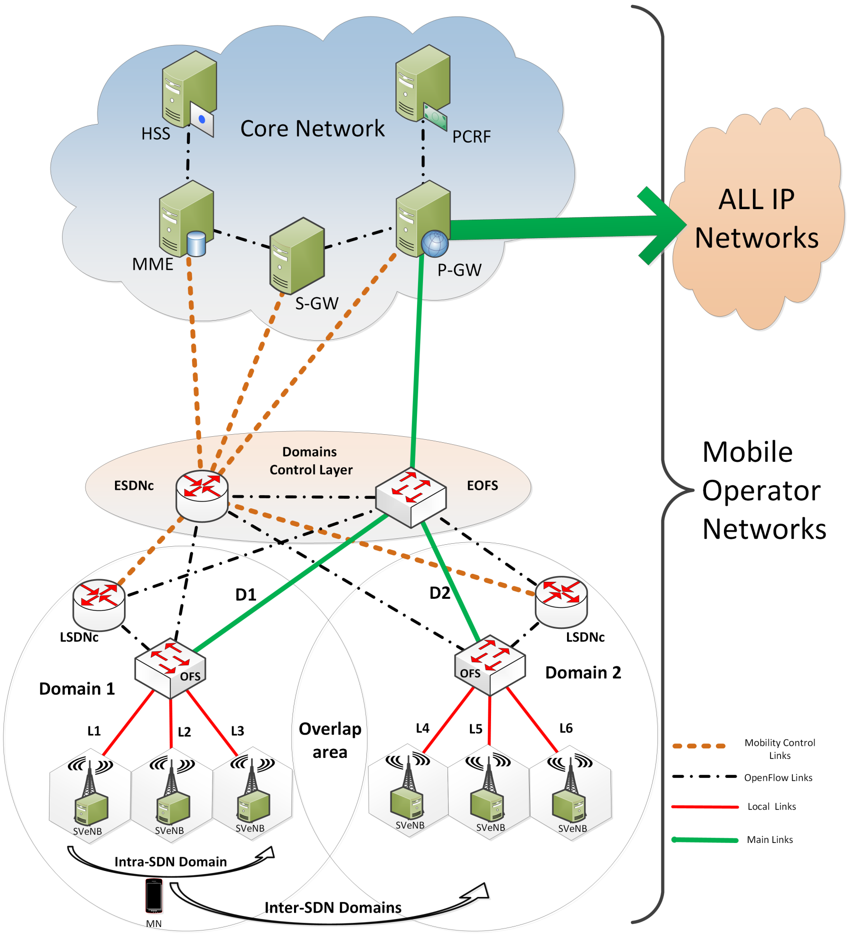


Fig. 4.3 The Proposed System Architecture

second device in this layer is EOFS which receives the data traffic from P-GW server of the core network. The EOFS tests the incoming packets with the entries of its flow tables to forward the flow to the destination target.

ESDNc is the central brain of forwarding decisions in the domains control layer. The responsibility of ESDNc is to make decisions to control the EOFS to forward data through specific main links ($D1$ or $D2$) which connect EOFS with each OFS in domain 1 or domain 2 respectively. Table 4.1 illustrates the contents of the lookup table of ESDNc such as main links, routing prefix ID, local links and T_Hs. i represents *ith* tag of the MN_i that belongs to link *ith* and N represents the number of users at each link.

Table 4.1 The ESDNc entries lookup table

Main Link	Local Link	Prefix ID	MN ID Tags
D1	L1	2001:0DB8:ACAD:0001::/64	$T_{H_{i_1},i=1...N_1}$
	L2	2001:0DB8:ACAD:0002::/64	$T_{H_{i_2},i=1...N_2}$
	L3	2001:0DB8:ACAD:0003::/64	$T_{H_{i_3},i=1...N_3}$
D2	L4	2001:0DB8:ACAD:0004::/64	$T_{H_{i_4},i=1...N_4}$
	L5	2001:0DB8:ACAD:0005::/64	$T_{H_{i_5},i=1...N_5}$
	L6	2001:0DB8:ACAD:0006::/64	$T_{H_{i_6},i=1...N_6}$

4.4.2 Domain Entities Task And Function

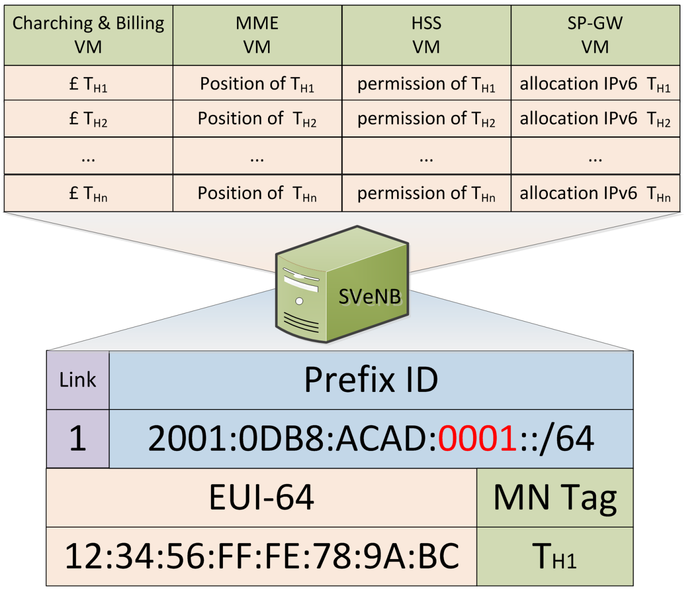
To illustrate the functions of each entity in the domain, we will discuss each entity tasks and what functions can be executed by it.

- SVeNB [63] is regarded as a macrocell base station. The SVeNB virtually serves as an eNB with a capability to host several VMs which mimic functionalities of mobile operator core network entities such as MME, S-GW, P-GW, etc. These entities are virtualized in multi VMs into SVeNBs. The virtual MME (vMME) which serves as a local vMME entity, virtual serving/packet (vSP-GW) which implements as local virtual serving/packet gateway and so on.

  The VMs partially perform the functions of the core network, due to the profiles of users were manipulated and updated in the core network and sent to the SVeNB. Therefore, the VMs can serve the local users that are covered by SVeNB without contacting the core network again when an MN tries to make a connection with another MN within the coverage area of an SVeNB [63].

  To move from SVeNB to another (i.e., from subnet to another) vMME, vSP-GW VMs play the primary role to bind IP address of an MN and determine its location. VMs of SVeNB can achieve these tasks. In other words, SVeNB can tie an MN IPv6 address (prefix and EUI-64) with T_H in the routing table which is built by vSP-GW in SVeNB as shown in Fig. 4.4.

- OFS provides packets forwarding within the domain. The domain could have more than one switch. We supposed using one switch for simplicity. OFS requests the forwarding decisions from the LSDNc to flow the packets to the target SVeNB. Then these decisions are cached in its



Charching & Billing VM	MME VM	HSS VM	SP-GW VM
£ T_{H1}	Position of T_{H1}	permission of T_{H1}	allocation IPv6 T_{H1}
£ T_{H2}	Position of T_{H2}	permission of T_{H2}	allocation IPv6 T_{H2}
...	...	...	...
£ T_{Hn}	Position of T_{Hn}	permission of T_{Hn}	allocation IPv6 T_{Hn}

Link	Prefix ID
1	2001:0DB8:ACAD:0001::/64

EUI-64	MN Tag
12:34:56:FF:FE:78:9A:BC	T_{H1}

Fig. 4.4 SVeNB VMs and Routing table Based on T_H

forward table for a given time at the forwarding device. The OFS contains one or more of flow tables, which consist of forwarding entries. These entries determine how packets are forwarded and processed according to the entries of a flow table which matches these packets. The typical entries of flow table are *(1) matching rules, or match fields* which contain information to be matched with those in the header of arrived packets, metadata, and ingress port. *(2) counters* which collect statistics such as the number of bytes, number of arrived packets and period of the flow for a certain flow. *(3) actions* which are applied a set of instructions on received packets to dictate how to forward the matching data [115].

- LSDNc can be considered as the main brain to make decisions to forwarding packets that are incoming to the domain. Additionally, it is responsible for making the decisions to the OFS to forward the data to a specific SVeNB which delivers the data to destination MN. These taken decisions are based on the lookup table that was saved and updated in LSDNc. The lookup table consists of SID, T_H, and link which represents the port ID that connect with the a specific SVeNB. Table 4.2 shows the entries of the lookup table. The subnet ID part (*16 bits with red color*) represents the local link topology of the mobile operator networks (domains).

Table 4.2 The LSDNc entries lookup table

Local Link	Prefix ID	MN ID Tags
L1	2001:0DB8:ACAD:0001::/64	$T_{H_{i_1}}, i{=}1 \ldots N_1$
L2	2001:0DB8:ACAD:0002::/64	$T_{H_{i_2}}, i{=}1 \ldots N_2$
L3	2001:0DB8:ACAD:0003::/64	$T_{H_{i_3}}, i{=}1 \ldots N_3$

The connection between the LSDNc and OFS utilizes OpenFlow protocol [116]. By using OpenFlow protocol, the LSDNc can add, remove, or update flow entries of the OFS flow tables to support the MN to receive packets when it moves amongst SVeNBs belong to a domain.

4.5 Mobility Management

The most robust feature of SDNc is that the ability to manage the mobility for each flow [117]. This feature enables the forwarding, load balancing and packets handover in both intra-domain and inter-domain. Moving an MN from an SVeNB to another, or from domain to another, needs that the new attachment point receives information about the MN from old SVeNB or from the MN itself which involves in the handover. Based on this information, a new binding table should be built. Fig. 4.3 shows the mobility of an MN into the intra-SDN domain and Inter-SDN domains respectively.

4.5.1 Inter-SDN Domains Mobility Management

The domains control layer is regarded as the first line to filter incoming packets from the core network due to the operation of checking which is done by the ESDNc and EOFS. This filtering process can be considered as packets sifting. Fig. 4.5 illustrates the proposed procedures which are implemented by EOFS and ESDNc, i.e., domains control layer. ESDNc can manage the horizontal handover (as defined in Section 4.6) amongst the domains. This approach begins when the MN enters overlap area, i.e., moves from the hosted domain to a new domain. Horizontal handover begins when the MN detects the signal of the visited SVeNB at the overlap area to register to the new SVeNB and sends its T_H. At the same time, the MN keep the connection with the previous SVeNB to send and receive data. The new SVeNB binds the T_H in its routing tables and triggers the T_H to LSDNc which in turn

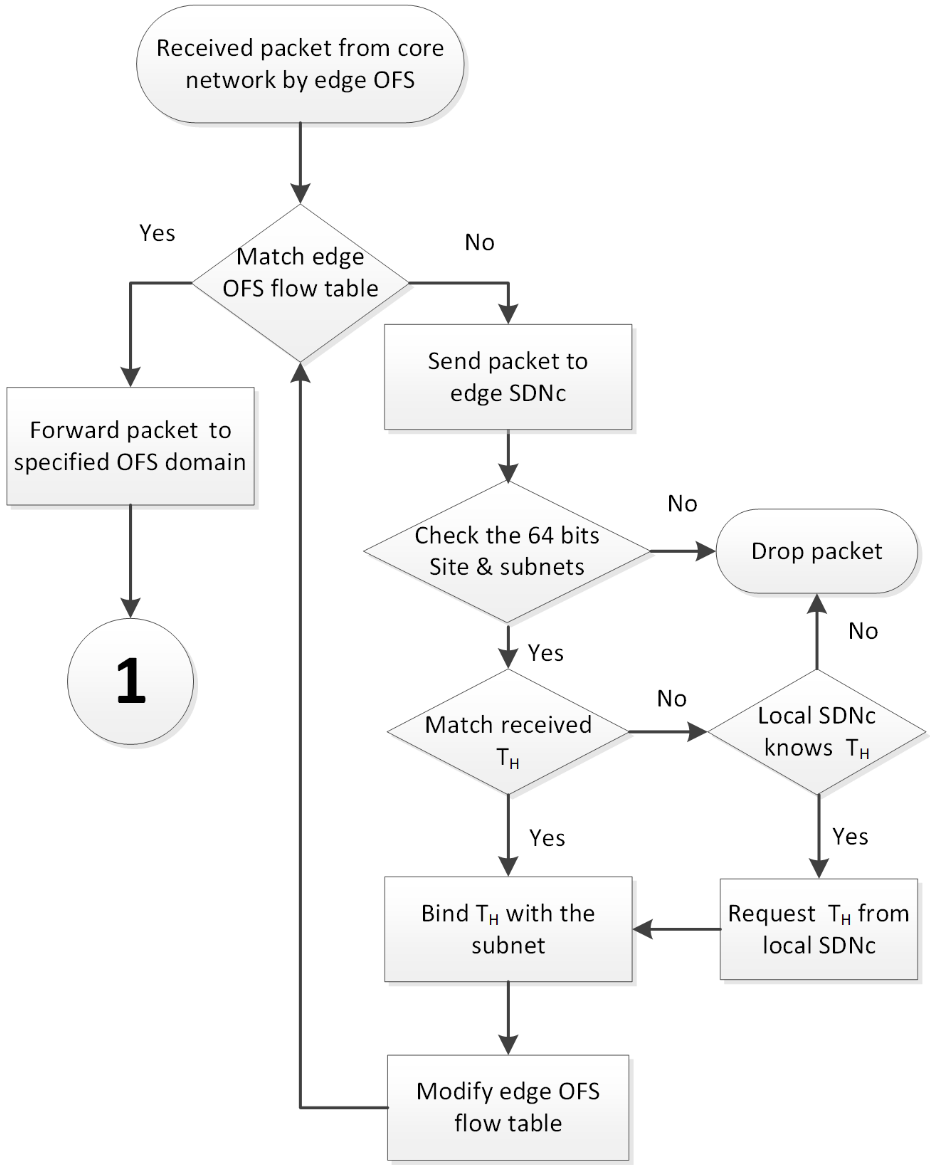

Received packet from core network by edge OFS
Match edge OFS flow table
Yes
No
Forward packet to specified OFS domain
Send packet to edge SDNc
1
Check the 64 bits Site & subnets
No
Drop packet
Yes
Match received T_H
No
Local SDNc knows T_H
No
Yes
Yes
Bind T_H with the subnet
Request T_H from local SDNc
Modify edge OFS flow table

Fig. 4.5 Flowchart of Forwarding Packet in Domains Control Layer

sends the T_H to ESDNc to modify forwarding tables of EOFS. The EOFS sends the last packet on the previous link until the EOFS executes the modification of the flow table.

4.5.2 Intra-SDN Domain Mobility Management

When an MN moves from one SVeNB to another within one domain, it should declare its T_H to the new SVeNB which requests the profile of that MN either from the old SVeNB or the mobile operator core network. The new SVeNB advertises the T_H of the MN to the LSDNc to bind with its prefix

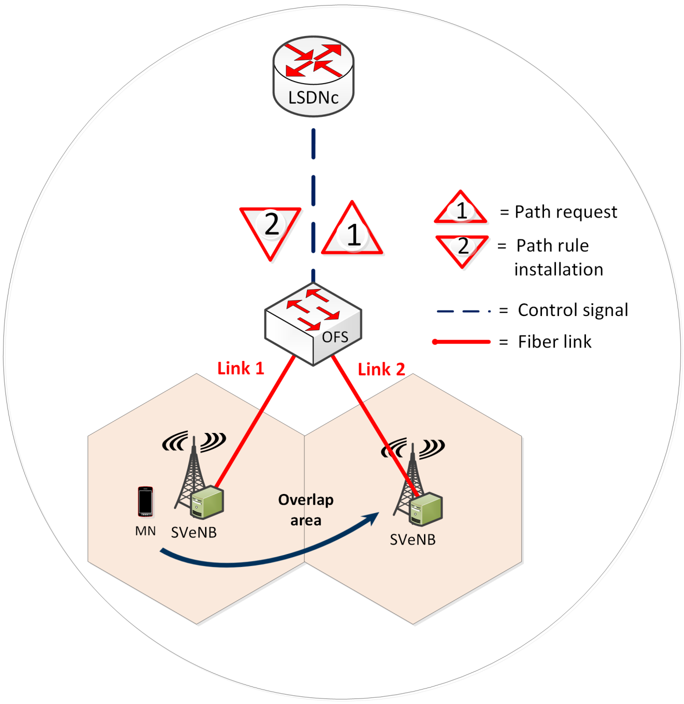


Fig. 4.6 Intra-SDN Domain Mobility

ID and link. Fig. 4.6 shows the mobility management by the LSDNc. The packet is received by the OFS which checks its flow table to forward that packet. If the OFS finds a match for that packet, it immediately forwards to the target SVeNB. If OFS does not find a match, then it sends that packet (step 1) to the LSDNc. The decision is replied by the LSDNc (step 2) whether modification the flow entries or dropping that packet.

- At the beginning, the MN generates its T_H via the operation aforementioned in Section 4.3. The MN requests a radio frequency bearer after it detects the signal of an SVeNB. The VMs which were installed into the SVeNB manipulates the session establishment. This process is known as access stratum connection. Meanwhile, the MN sends its T_H to SVeNB that covers that MN. A standout the majority advantages of using IPv6 is its capability with auto-configuration addressing. An MN can configure its IPv6 address according to the link-local prefix ID for each interface. This procedure is known as a stateless auto-configuration IPv6 creation which depends on IID of an MN EUI-64 (based on MAC address) and link prefix to form a global or local address [14]. An installed VM into SVeNB binds T_H with the IPv6 address of the MN in forwarding or routing table which is used by SVeNB to deliver the packets to the MN. Also, this table is used by vMME to locate the position of that MN. Fig. 4.7 shows the messages between the MN and SVeNB to establish the connection.

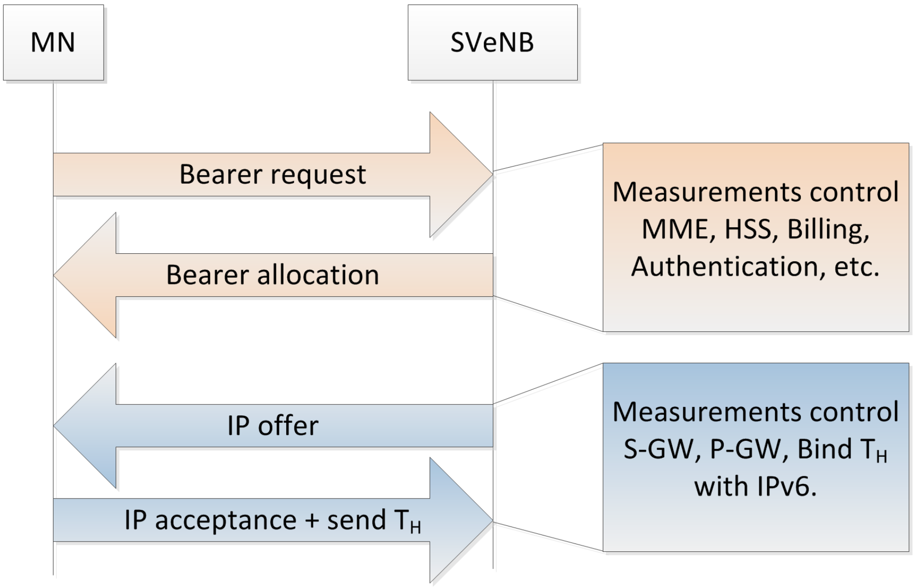

MN
SVeNB
Bearer request
Bearer allocation
Measurements control MME, HSS, Billing, Authentication, etc.
IP offer
IP acceptance + send T_H
Measurements control S-GW, P-GW, Bind T_H with IPv6.

Fig. 4.7 The Messages Between MN and SVeNB

The necessary measurements such as authentication, mobility, user service permissions, etc., these measurements Should be accomplished by installed VMs into SVeNB.

- After the MN is registered by an SVeNB. The T_H is directly sent to the LSDNc to update its lookup table. Subsequently, the local SDNc sends the new rules and actions to the OFS which updates its forwarding table. At the same time, LSDNc triggers the received T_H to the ESDNc to know the new locator (subnet) of the MN. Fig. 4.4 illustrates the VMs with user profiles and routing table based on T_H. The SVeNB plays the significant role of connecting all users that are under the coverage area of that SVeNB. Moreover, it receives packets from the source through the OFS to deliver that packets to the destination MN. The vMME manages the functionality of MN mobility within the coverage area, i.e., it manages the local mobility of all users under the tent of the SVeNB.

- The LSDNc receives packets that should contain SVeNB prefix ID (locator) and the MN host ID (T_H) of attached SVeNB. These packets are used by LSDNc to update its lookup table and maintain the mobility of that MN. In other words, the locator represents the subnetting topology and geographic location of SVeNB which is already known by the LSDNc. While T_H represents the position of an MN that should be bound with that locator to be known by LSDNc. This location and position awareness can consider as domain mobility management, due to determination the topology identifier (prefix ID) and the MN identifier (T_H) which are represented as the locator and position of the MN respectively. Algorithm 1 illustrates the procedures that are taken by LSDNc to make decisions and update its lookup table. Fig. 4.8 illustrates the proposed checking and processing in SDN domain.

4.6 Handover Procedure

The mobility management emerges to solve the problems of roaming the MN among the wireless mobile networks. Mobility management preserves the continuity of MN connection when the MN alters its attachment point to a new network; this is called handover management. Besides, mobility management enables the MN to receive packets from serving networks at different access points of attachment, this is known as location management. [118].

There are two types of handover, vertical handover which means that the MN moves among different technologies of wireless access points. For example, WiFi, WiMAX, LTE, etc. Vertical handover can be done in the same geographic area has a variety of wireless coverage connectivity. The second type is horizontal handover which refers to the MN when it moves with the same technology in different geographic areas.

Each IPv6 address carries network identifier (prefix ID) which consists of 64 bits of the IP address and host identifier or interface identifier (IID) which consists of the other 64 bits of the IPv6 address. Prefix ID has topological importance due to the routers use prefix ID to forward the packets among different networks, i.e., at the network layer. While IID is topologically important at the target subnet to deliver data to the MN belong to that subnet. In our proposal the T_H is equivalent to the IID indicates the position of the MN at a specified subnet [119].

The system proposes using T_H within domains control layer and the domains themselves only. Since the connection between an MN and its sender node uses the standard IPv6 to keep receiving and sending packets from/to backbone networks. The CN does not be aware of the MN location.

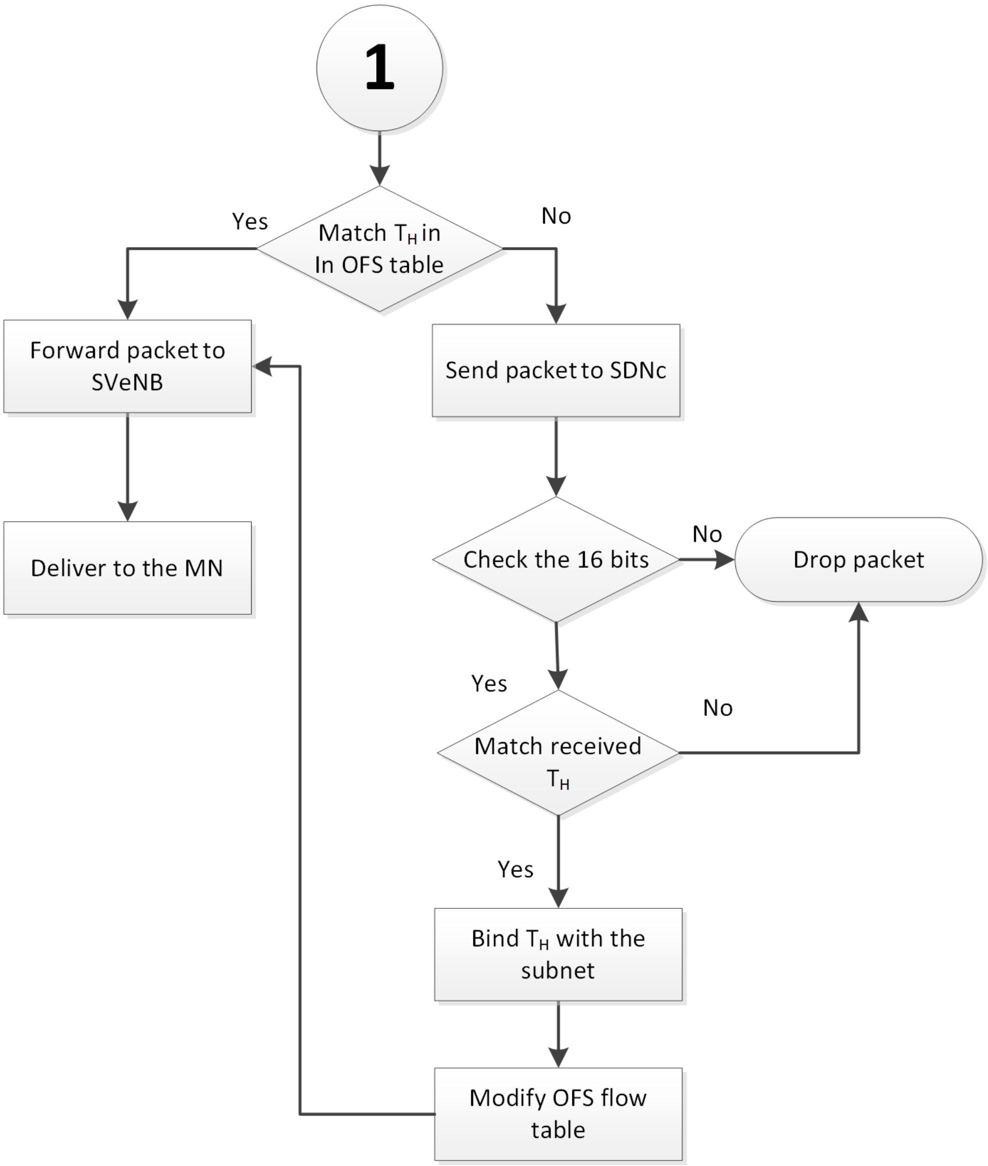


Fig. 4.8 Flowchart of Forwarding Packet in SDN Domain

Consequently, the connection is continuous and uninterrupted that leads to seamless and almost zero delay handover. Fig. 4.9 the links type according to use the standard or non-standard IPv6 routing schemes. The handover between the domains starts at ESDNc after receiving information about the new binding of T_H and the visited subnet from LSDNc. ESDNc makes modifications and changes on its routing flow tables. These modifications and changes send to EOFS which changes the exit link from $D1$ to $D2$ to forward packets as shown in Fig. 4.9. That happens when the MN enters the overlap area and after registering to the visited domain.

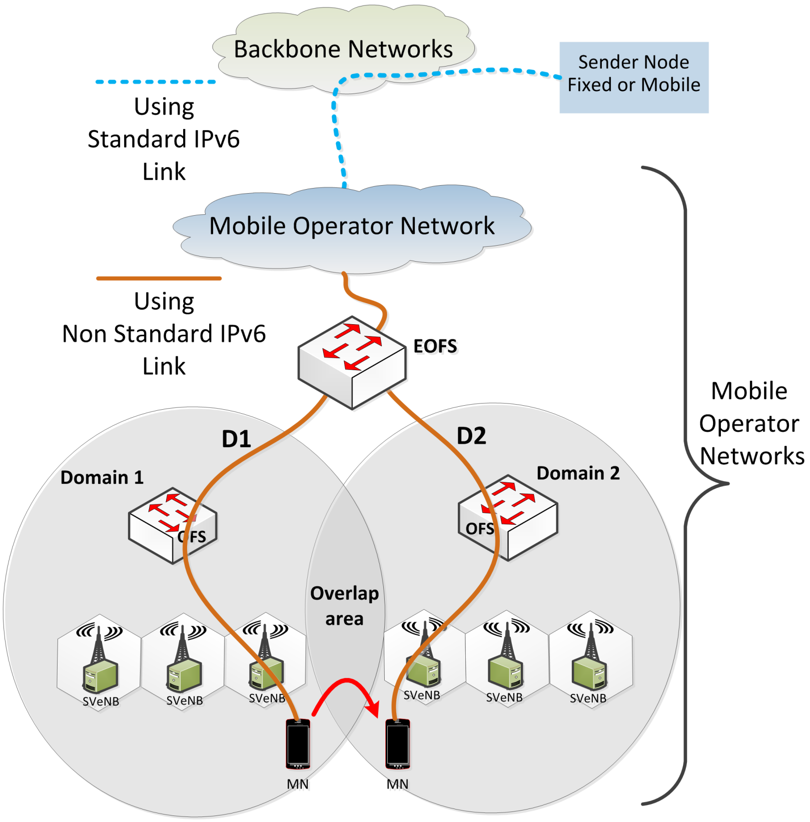


Fig. 4.9 The Link Type Based on Using Standard or Non Standard IPv6

4.6.1 Handover Delay

Currently, packets pass via the HA and FA to deliver data between the MN and sender node. That needs to more procedures to create tunnelling between the MN and sender node to keep the connection continuity. The most packets handover delay happens due to the processing procedures to change the route or flow of packets. If we consider in C-RAN systems, there are many physical servers each one in charge to achieve a specific task. Some of these servers located in one geographic area, and some other located away from each other. Moreover, some servers depend on the decisions of the other to complete its task. In both cases, all servers should process incoming data also process the preparation to send that data. Furthermore, queuing delay which depends on the amount of data on the link that can transfer data between two points, also depends on the hardware specifications of the servers.

4.6.2 Packets Path Decision Delay

The proposed system suggests using non-standard IPv6 routing scheme to forward packets through the system based on separation of the CP from DP. This feature of separation is supported by using the OpenFlow protocol [117]. The MN declares its T_H to SVeNB after acquiring prefix ID from the SVeNB which updates and binds the information of that MN. At the same time, SVeNB advertises T_H to the LSDNc to updates its lookup table and sends the modification entries of the flow table to the OFS.

There are two probabilities for the received packet by the LSDNc. Firstly, the received packet already has been bound with the subnet and the link in the lookup table. In that case, the decisions are sent to the OFS to forward that packet and keep that *matching rules* and *applying actions* for all packets which match that rules. Secondly, the packet is received by LSDNc for the first time; it checks the subnet of the prefix ID *(only the 16 bits)*. This checking has done by LSDNc to know if this packet belongs to one of its subnets or not. If the answer is yes, the LSDNc scans its lookup table to see whether T_H within that subnet or not. If the answer is no, then there are two likelihoods, the first one is that the LSDNc sends a request to SVeNB and ESDNc about that T_H to make a decision.

The second likelihood is that dropping all packets for which information is not found about their T_H at the LSDNc, SVeNB or ESDNc. In other words, the OFS drops all packets that are not matching or that are not known their T_H or subnet by SVeNB, ESDNc, and LSDNc. Algorithm 1 shows the procedures have adopted to forward packets within the proposed system.

4.6.3 Processing and Queuing Delay

To calculate the packets handover delay we need to determine the process and queue delays with each server. Suppose every server achieves one task. Consider the queuing of the proposed system is M/M/1 with Poisson process. Let § is the number of servers, μ is the packet transmission rate of the control messages and λ is the Poisson arrival process rate (packet/sec) at each server which provides a traffic load as,

$$\rho_i = \frac{\lambda_i}{\mu_i}, \qquad i = 1, 2, ..., \S \tag{4.1}$$

Algorithm 1: ESDNc and LSDNc Algorithm to Make a Flow Decision

```
1  Packets received from core network
2  if Received packet prefix (64 bits) match then
3      if $T_H$ bound with main link then
4          Send packet through the specific main link
5      else
6          Request $T_H$ from LSDNc
7          Update tables of ESDNc and EOFS
8          if Subnet (16 bits) match then
9              if $T_H$ belongs to a subnet then
10                 Send to SVeNB belongs to that subnet
11                 Deliver to the MN
12             else
13                 Request $T_H$ from SVeNB
14                 Update tables of LSDNc and OFS
15                 Send trigger of $T_H$ to ESDNc
16                 Go to step 7
17             end
18         else
19             Drop packet
20         end
21     end
22 end
```

Where, ρ_i and λ_i are the utilization factor and arrival rate of *ith* server respectively [120]. The total delay of expected queuing equals to the summation of expected queues at every server. So, it is expressed as,

$$E[\chi] = \sum_{i=1}^{\S} E_i \left[\frac{1}{\mu_i} \right] \tag{4.2}$$

Where χ is the service delay of a server and equals to $1/\mu$. Based on first-come-first-serve, inter-arrival times, service times are independent, and by using Markov chain then the probability ($Prob_{cm}$) of control message packets be in the queue is,

$$Prob_{cm} = \rho^B(1-\rho) \tag{4.3}$$

Where B is the number of the control messages that transferred in a channel, and $\rho = \lambda/\mu$, From this we can get $B = \rho/(1-\rho)$. So the delay for each server is:

$$D_{qs} = \frac{B}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{(\mu-\lambda)} \tag{4.4}$$

Then the overall delay due to queuing in C-RAN is given as in Equation 4.5 , where § is the number of servers,

$$D_{qT_t} = \sum_{i=1}^{\S} \frac{1}{(\mu_i - \lambda_i)} \tag{4.5}$$

Whereas for the proposed scheme we adopted the queue system M/M/m [121], and the equation of the total queuing delay is,

$$D_{qProp_t} = \sum_{i=1}^{Ph} \frac{1}{(m\mu_i - \lambda_i)} \tag{4.6}$$

Ph is the number of physical servers in proposed architecture, and *m* is the number of performed tasks simultaneously by the ith server .

The queuing delay depends on several physical parameters such as transmission line capability and the specifications of the server components like memory size, Central Processing Unit (CPU), etc. We named the processing delay for all components cause delays. Assume the packet length is L_p, the machine word size is W_m, the arrived word size is W_a, the number of the packet in each control message is P_{cm}, CPU architecture of the server (e.g.,32 or 64 bit) is CPU_x. Moreover, the lookup delay of memory access is assumed almost 100 nsec [63]. The processing delay equation can be modified and written as,

$$D_{ps} = 100\frac{W_a}{W_s} \times \left[log_{sys} P_{cm} + \frac{L_p}{CPU_x} \right] \tag{4.7}$$

Equation 4.7 is used for one physical server. The total delay of the servers in the traditional system equals to the summation of delays of the servers that have been involved in making the decision for packets routing.

$$D_{pt} = \sum_{i=1}^{\S} \sum_{u=1}^{U} 100\frac{W_{aiu}}{W_{si}} \times \left[log_{sys} P_{cmiu} + \frac{L_{p_u}}{CPU_{xi}} \right] \tag{4.8}$$

Where § is the number of servers, U is the number of control messages of uth user to the specific ith server to be processed and sys is the system constant.

In our proposed system, the number of servers is much lower than in C-RAN system due to the VMs which were installed at SVeNBs. These SVeNBs perform the duties of physical servers of the core network to sustain CP of any connection. Moreover, using SDNc and OFS give the ability to decrease the number of control messages of each packet flows. These essential pros can be noticed in enhancing packets handover process and reducing the packets handover delay. Fig. 4.10 shows the results of the packets loss probability by increasing the number of users. It is clear that the traditional system presents larger values of packets loss than SDN system. In traditional scheme, the packets forwarding process on the network devices such as queuing, processing, decapsulation, encapsulation, etc. are executed on each packet enters those devices. While in SDN system the packets are sent as a flow. Each flow consists of a group of packets. These flows are directed based on the forwarding process which is done only on the first packet of each flow. In other words, all the packets of a flow track the first packet. That led to keep the probability of packets loss in SDN system almost in range of one-seventh of that in the traditional system. Fig. 4.11 shows the results of the received packets during the mobility of the MN. As shown in Fig. 4.11 as the MN moves slowly; the performance of

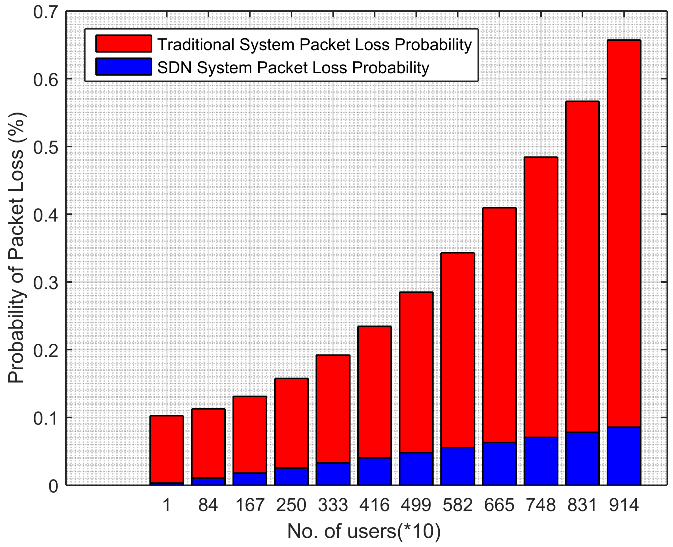


Fig. 4.10 The Probability of Packets Loss

both systems is high, due to both systems considered the MN as a fixed node. Therefore, the MN can receive packets with the minimum probability of packets loss. That reflects the packets did not need to change or modify its route.

With increasing the speed of the MN, the proposed system recorded higher performance of than the traditional system. Moreover, the intra-SDN domain handover scenario performed higher receiving of packets than the inter-SDN domains handover scenario. That because of the LSDNc handled the packets within the domain. While in case of inter-SDN domains the performance was slightly less than the intra-SDN domain, due to the packets were directed by ESDNc and LSDNc. The performance of both intra-SDN and inter-SDN domains are higher than the traditional system at increasing the MN speed. Considering more than one network device governs the packets forwarding in the traditional system, due to each device makes its forwarding decision (each device gathered CP and DP into its structure).

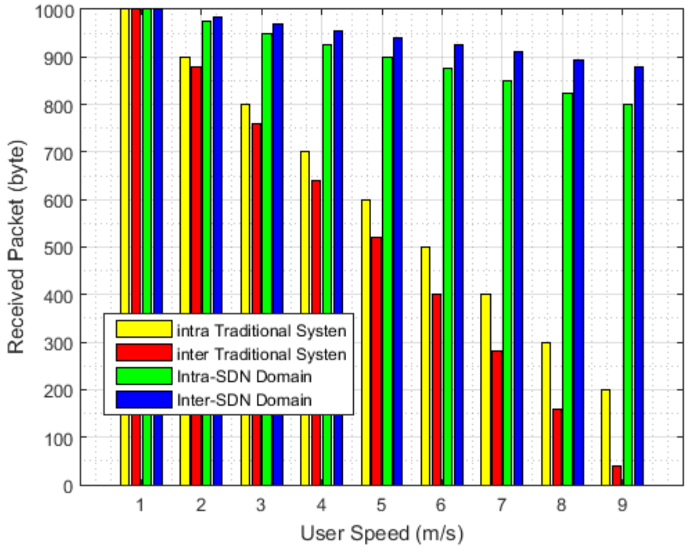


Fig. 4.11 The Received Packets During MN Movement

Table 4.3 Comparison Between The Proposed Scheme And Other Protocols Scheme

Protocols	HIP	Shim6	LISP-MN-Local	MOFI	LISP-AR-DMC	DHT-MAP	Proposed Scheme
Centric Type	Host-based	Host-based	Host-based	Host-based	Network-based	Network-based	Network-based
Mapping Type	Centralized	Centralized	Centralized	Distributed	Distributed	Distributed	Distributed
Management Manner	Rendezvous Server	DNS Server	LMS	LMCs	LMS	Rendezvous	SDNc
Decoupling CP & DP	No	No	No	No	No	No	Yes
Deployment Cost	High	High	High	High	Low	Low	Low
Packets Forwarding	Tunneling	Tunneling	Tunneling	Tunneling	Tunneling	Tunneling	Based-Fow Table
Direct ID Forwarding	No	No	No	No	No	No	Based-T_H
Dispatch Path	Routing	Routing	Routing	Routing	Routing	Routing	Flow Forwarding

4.7 Comparison With Other Schemes

The points of similarity and difference between the proposed scheme and other schemes are summarised in Table 4.3. We can notice from the Table 4.3 that the unique and shared points of the proposed and with the other schemes.

Host-based requires to amend the protocol stack of a host, so it leads to more cost and deployment problem [112]. Whereas, network-based does not need much modification in a protocol stack. Therefore, it is more acceptable and cost-efficient of engaging with SDN environments.

The centralized management suffers from traffic burden, single point of failure and centralised mapping. To overcome these obstacles either by adding more devices with high or super specifications for the server to contribute in data processing and this leads to high cost or distributing the services amongst more than one device with reasonable specifications to afford the cost. The second choice can be realised by utilizing SDN technique to distribute the tasks between two devices with SDN-enabled technique with reasonable specifications. Data processing delay decreases due to the jobs are treated in parallel at the same time. Particularly, when separate the DP from the CP into different devices (i.e., flow forwarding and flow decision maker respectively) [106].

The proposed system is unique to these features which all are grouped in ours and not found in the others by:

1. Depending on decoupling the CP from the DP to forward packets.
2. Managing mobility per flow for an MNs.
3. Using direct forwarding to the MN based on the generated T_H.
4. Implementing flow forwarding instead of routing or switching.

These are the substantial differences between our proposed scheme and the other schemes.

4.8 Simulation and Performance Evaluation

4.8.1 Mininet Simulator

To implement and evaluate our proposed network, Mininet simulation has been used. Mininet is a network emulator to simulate virtually the functionalities of the network devices (servers, routers, switches, hosts, and links). Mininet is a significant means to work on underlying the open source software such as SDN, NFV, and networks virtualization. Moreover, it can be used to design an actual virtual network similar to the real functioning of the network components but operate on one physical or virtual machine. Mininet allows creating custom topologies and gives the ability to create and configure controllers, switches, and hosts through:

- Interactive User Interface (IUI).

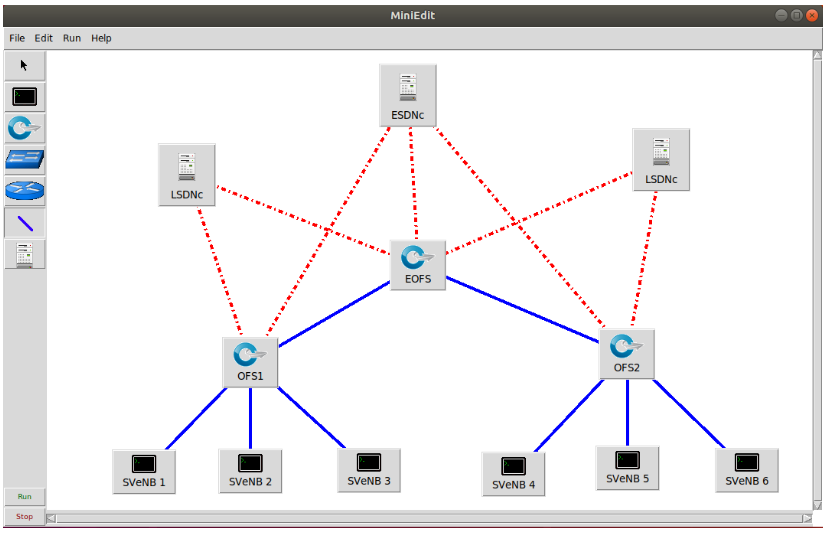


Fig. 4.12 The Mininet Proposed Scenario Setup

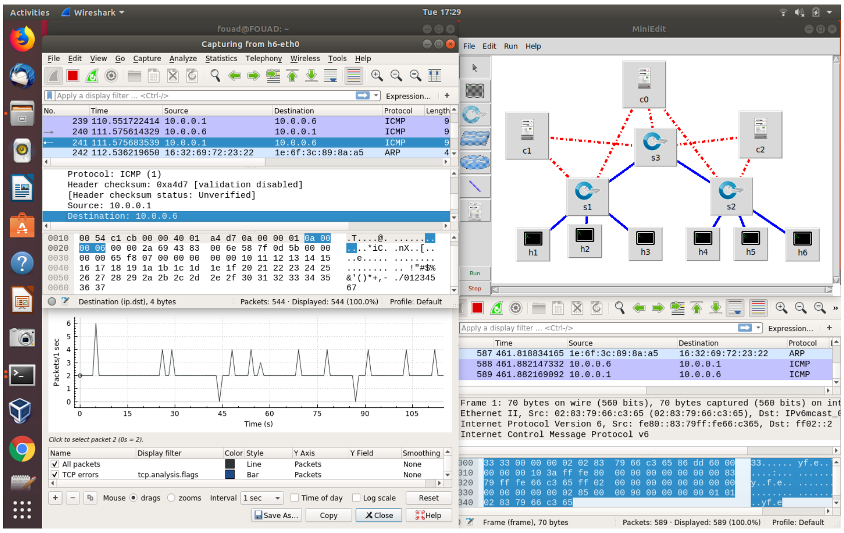
- Command Line Interface (CLI).

- Programming Languages such as Java, Python, etc.

The Mininet simulator is used to implement our SDN network system. The simulation scenario consisted of three OpenFlow enabled switches (EOFS, OFS1, OFS2), and two sets, each set with three hosts. The first set connected with OFS1 and the second set connected with OFS2. The hosts represent the SVeNBs to mimic the stationary parts of the mobile network. The Python language has been used to configure the APIs of the simulation scenario. Figs. 4.12 and 4.13 present the setup of the proposed network and execution under Linux operating system.

4.8.2 Performance Evaluation

The performance evaluation comparisons of the packets handover, addresses mapping, links switch and processing delays of CP consideration for both traditional (C-RAN) system and our proposed system. The MATLAB platform was used to collect the datasets which were prepared and pre-processed for implementing in this system. The MATLAB has been used to evaluate the performance of the proposed algorithms; the relevant simulation parameters have recorded in Table 4.4. The extracted data has been injected to evaluate the performance measurements of our system behaviour.



```
No.  Time            Source              Destination         Protocol Length
 239 110.551722414 10.0.0.1            10.0.0.6            ICMP     9
 240 111.575614329 10.0.0.6            10.0.0.1            ICMP     9
 241 111.575683539 10.0.0.1            10.0.0.6            ICMP     9
 242 112.536219650 16:32:69:72:23:22   1e:6f:3c:89:8a:a5   ARP      4

    Protocol: ICMP (1)
    Header checksum: 0xa4d7 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.0.0.1
    Destination: 10.0.0.6

0010  00 54 c1 cb 00 00 40 01  a4 d7 0a 00 00 01 0a 00   .T....@. ........
0020  00 06 00 00 2a 69 43 83  00 6e 58 7f 0d 5b 00 00   ....*iC. .nX..[..
0030  00 00 65 f8 07 00 00 00  00 00 10 11 12 13 14 15   ..e..... ........
0040  16 17 18 19 1a 1b 1c 1d  1e 1f 20 21 22 23 24 25   ........ .. !"#$%
0050  26 27 28 29 2a 2b 2c 2d  2e 2f 30 31 32 33 34 35   &'()*+,- ./012345
0060  36 37                                              67

Destination (ip.dst), 4 bytes    Packets: 544 · Displayed: 544 (100.0%)    Profile: Default
```

Click to select packet 2 (0s = 2).

Name	Display filter	Color	Style	Y Axis	Y Field	Smoothing
☑ All packets		■	Line	Packets		None
☑ TCP errors	tcp.analysis.flags	■	Bar	Packets		None

Mouse ● drags ○ zooms Interval 1 sec □ Time of day □ Log scale

```
     Time            Source              Destination         Protocol
 587 461.818834165 1e:6f:3c:89:8a:a5   16:32:69:72:23:22   ARP
 588 461.882147332 10.0.0.6            10.0.0.1            ICMP
 589 461.882169092 10.0.0.1            10.0.0.6            ICMP

Frame 1: 70 bytes on wire (560 bits), 70 bytes captured (560 bits) on int
Ethernet II, Src: 02:83:79:66:c3:65 (02:83:79:66:c3:65), Dst: IPv6mcast_
Internet Protocol Version 6, Src: fe80::83:79ff:fe66:c365, Dst: ff02::2
Internet Control Message Protocol v6

000  33 33 00 00 00 02 02 83  79 66 c3 65 86 dd 60 00   33...... yf.e..
010  00 00 00 10 3a ff fe 80  00 00 00 00 00 00 00 83   ....:... ......
020  79 ff fe 66 c3 65 ff 02  00 00 00 00 00 00 00 00   y..f.e.. ......
030  00 00 00 00 00 02 85 00  00 90 00 00 00 00 01 01   ........ ......
040  02 83 79 66 c3 65                                  ..yf.e

Frame (frame), 70 bytes    Packets: 589 · Displayed: 589 (100.0%)    Profile: Default
```

Fig. 4.13 The Proposed Scenario Execution

Fig. 4.14 shows the delay time difference between the proposed scheme and traditional schemes to setup flow path connections between the MN and the sender node. Fig. 4.14 declares that the traditional schemes necessitated more time in a queue to determine the path for an arrived packets. This queuing is repeated for each packet. While in our proposal flow path determination demanded much less time due to making decisions have been achieved in one or two servers each one with multiple VMs.

For example, the figure shows that for 200 packets, the delay is 0.182 ms in the proposed scheme, while for the same number of packets in the traditional scheme is 0.2 ms. The delay time at 1000 Packets is 0.62 ms, while for 1000 packets in traditional system scheme is 1.4 ms. It is clear that the delay time is more in the traditional system than the delay in the proposed system; due to the parallel processing that has been applied by the VMs on the arrived packets in our proposal.

Fig. 4.15 illustrates the addresses mapping delay versus the control messages. Each control message contains many packets. In the traditional system, some control messages can be considered as *MasterMessages* are generated and sent by a server to another server as complementary control message as *SlaveMessage*. Therefore, the line graph of the traditional system is exponential, and the delay grows with increasing the number of control messages.

Whereas, in the proposed system the same messages *MasterMessage* and *SlaveMessage* are processed in parallel into the same physical server through the VMs. This approach leads to decrease

Table 4.4 The System Parameters

Parameters	Value
No. of Users	10k
No. of Packets	1000
User Speed	Up to 9 m/sec
Packet Size	1522 Byte
Max. Control Messages	120 message
Min. Control Message Size	50 Byte
Delay Per Link	0.1 msec
Processing Delay	0.05 msec
No. of Re-directions	8
OFS Modifying Delay	0.005 msec
SDN Modifying Delay	0.001 msec
Virtual S/P-GW Delay	0.001 msec
Virtual MME Delay	0.001 msec
RF Intra Registration	1 sec
RF Inter Registration	2 sec
Simulation Rounds	12000

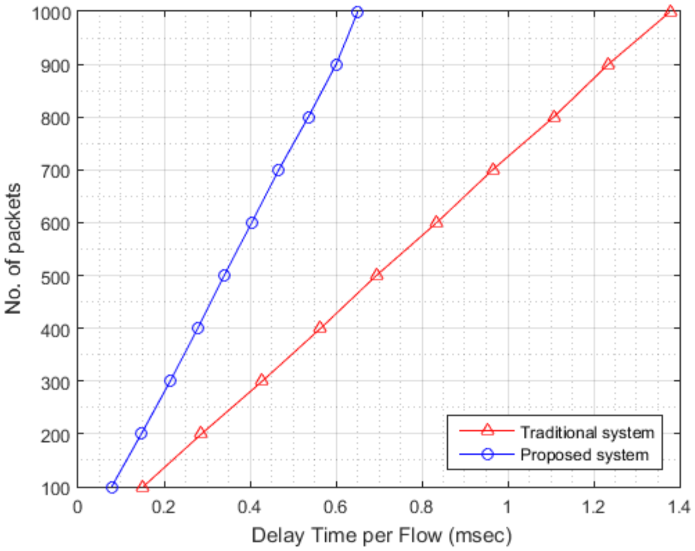


Fig. 4.14 The Delay Time Comparison Between SDN And Traditional of Forwarding And Routing Schemes Respectively

the required time to build of the addresses mapping tables. Fig. 4.15 shows the SDN environment can build its addresses mapping tables in less than one-tenth of the time that is needed by the traditional environment to build its addresses tables for the same number of control messages and number of users.

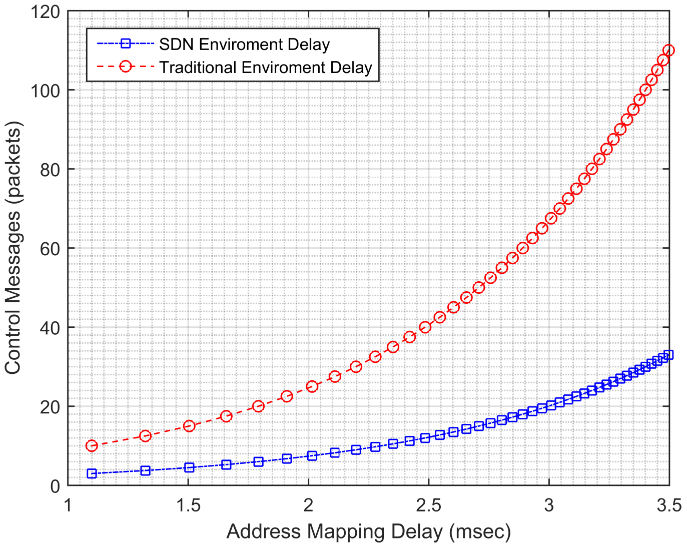


Fig. 4.15 The Delay Time to Create Addresses Mapping Tables

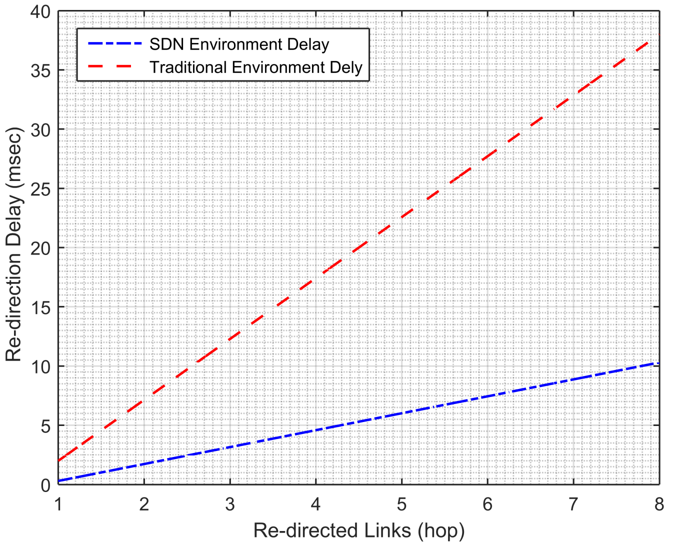


Fig. 4.16 The Delay of Links Switch

Fig. 4.16 represents the delay of a flow (packets) re-direction (number of hops) that should the packets pass through them to reach the target MN. As shown in the figure, the SDN environment needed almost 12.5% of control messages that were required by traditional systems. That because the link switch mechanism depends on tagging which was used by the proposed system.

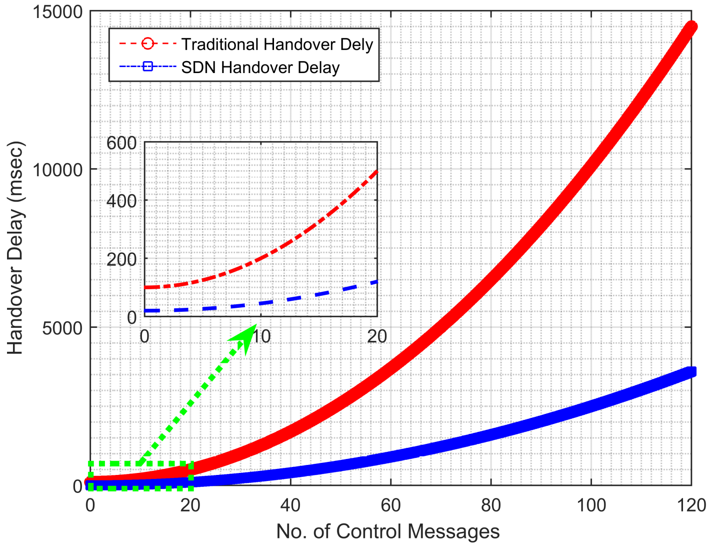


Fig. 4.17 The Overall Delay of Handover And Required Control Messages

In the instant of receiving T_H, the EDSNc modifies its lookup table then sends the amended flow path to the EOFS to redirect the packets flow from $D1$ to $D2$ which connect each OFS in domain 1 and domain 2 respectively, as shown in Fig. 4.16. At this time, LSDNc had been updating the flow table of OFS which subjects to the LSDNc and ESDNc administration.

Packets handover delay plays a vital role in the session continuity of an MN connection. Registration, getting the CoA and tunnelling are the main parameters for the packets handover delay. Our proposal shows seamless and extremely soft handover. From Fig. 4.17 we can see that the packets handover in the proposed system needed fewer control messages to make decisions and change the flow path of data packets. While in the traditional system that took more processing and time, thus led to losing packets during the handover.

Fig. 4.18 shows the received data during packets handover procedure in both traditional and proposed systems. From Fig. 4.18, we can see that the proposed system kept the average of received data almost at the high level during the packets handover procedure. Whereas, the received data drops to the lower level in the traditional system through the packets handover. The results as mentioned earlier were based on the assumption of an MN moves at the same speed in both states (proposed and traditional systems). Moreover, the maximum values of received packets in the proposed network were almost 5.98 kb (when SDN network used the T_H) and 5.7 kb (when SDN did not use the T_H)

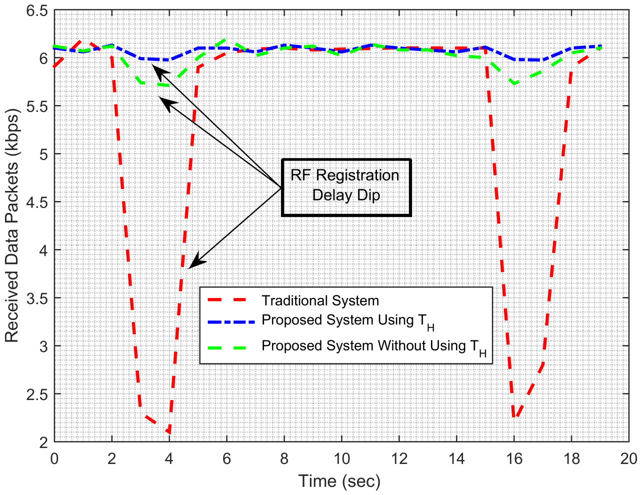


Fig. 4.18 The Received Data During The Handover Procedure

of 6.2 kb of unbuffered transmitted live stream respectively. In contrast, the minimum values of the received packets in the traditional system were around 2.25 kb of 6.2 kb of the unbuffered transmitted live stream. That means an MN can receive 96.4% and 91.9% of the packets that have transmitted during the handover process period in the SDN network based on the proposed packets forwarding and re-directing mechanism with using T_H and without using the T_H respectively. Whereas, by the traditional packets routing mechanism the received packets percentage is 36.3%. In other words, the proposed scheme can retrieve almost three times of that lost packets in the traditional system with neglecting RF registration delay time.

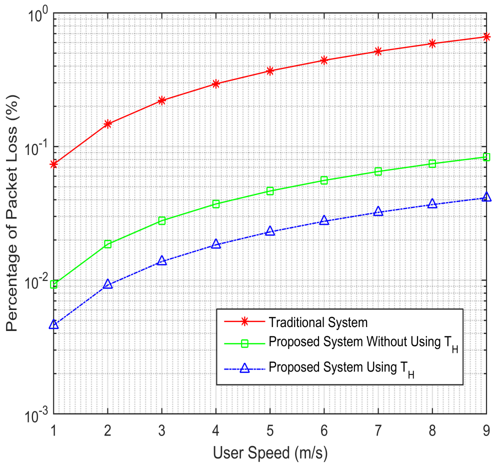


Fig. 4.19 Percentage of Average Packets Loss During Handover Process

Fig. 4.19 indicates the percentage of average packets loss against the MN speed. As expected, the SDN networks overcome the conventional networks in reducing the values of packets loss for unbuffered streams. This reduction in lost packets rate is due to the decrease in the required processing time to forward and re-direct packets into the SDN network (CP messages exchange). The performance of the SDN network has been enhanced by using the T_H, as shown in the Fig. 4.19, where the lowest value of lost packets was around 4% through using the proposed scheme with supporting of T_H and almost 8% without using the T_H in SDN network, while the lost packets value was nearly 34% in conventional network scheme.

4.9 Chapter Summary

In this chapter, a novel idea to generate a tag as MN identity from E.164 standard numbering and MAC address has been presented. Based on the uniqueness of E.164 numbering and MAC which are processed together to generate the MN tag (T_H). The T_H is used to handle the packets inside the domains. Also, it is used in seamless packets handover mechanism. By taking advantage of the SDN via separation the control plane and data plane. This decoupling is a suitable candidate to exploit it in our proposed system which uses SDN and other virtualization technologies. The requirements of 5G for future mobile communications urged us to think in a novel packets delivery during the packets handover to keep real continuity in connection to an MN. The proposed system and traditional system have been evaluated and simulated by MATLAB and Mininet platforms. Applying Smart virtualization architecture in mobile communication networks as a paradigm can impact, in general, the future of mobile communication networks. In this chapter, This chapter put forward a novel proposed system of smart virtualization for packets delivery, mobility management, and handover procedure comes down to the network-based. SDN and its integral OpenFlow protocol are used to separate the CP from the DP of network flow. This separation enables the mobile operator to control the infrastructure, reduce the operational and capital costs, and fulfill horizontal packets handover optimization. SDN can achieve the same duties and tasks that were accomplished by many physical devices can be performed and implemented by virtual network environments. SDN is perfect to simplify the management of IPv6 due to the potential of IPv6 such as the vast address space and the stateless auto-configuration.

Moreover, the IPv6 is not only used in mobile communications for routing purposes, but it can be accepted as a locator identifier and host identifier. These concepts are utilized by the proposed system which separates IPv6 into prefix ID and IID which are equivalent to locator ID and host ID respectively Our proposed system suggested an approach to generate host tag that to be employed as an indicator of MN movement between subnetworks. Consequently, horizontal packets handover can be achieved seamlessly with zero packets loss and extremely minimum delay time. All advantages aforementioned meets the 5G mobile networks for future mobile communications.

Chapter 5

Proactive Forwarding for High Data Rate in HST Networks

5.1 Introduction

Always being available and connected at any time, anywhere on the Internet by mobile or fixed devices is a feature of modern life. Increasingly, the development of application services is leading to need deliver large amounts of data without delay or interruption through online connection during user movement. To achieve this, an efficient network is required to handle the transferred data. When associating moving targets to a network, this is more difficult than for stationary, because their changing position and location for both geographic and topology of accessing attachment points of that network. Since packets forwarding amongst network devices depends on IP addresses, several protocols have been proposed to tackle packet loss and long handover latency through rapid acquiring the IP address of a mobile node. Such protocols include the Mobile IPv6 (MIPv6) [122], Fast Mobile IPv6 (FMIPv6) [123], Hierarchical Mobile IPv6 (HMIPv6) [124], Proxy Mobile IPv6 (PMIPv6) [125], and Fast Proxy Mobile IPv6 (FPMIPv6) [126]. Despite these protocols having been developed and provided significant advances in mobility management, they suffer from many obstacles such as long handover delay time, signaling overhead, and a high rate of packet loss [127]. These protocols serve as mobility management ones with the ability to handle such management through the IP layer. Moreover, they have provided significant features for mobility management at low and medium speed of moving targets[128]. However, the mobility management issue for high-speed moving targets is too difficult for the existing protocols and needs creative solutions to tackle this context. As a High-Speed Train (HST) moves from one AP to another, it requires disconnecting from the current AP (APc) and connecting to the next AP (APn) to sustain its session connectivity to the Internet, a process known as handover. The handover delay time is known as the timespan that is required by network devices to complete handover handling. Within the handover process, the HST is separated from the network for a while. Hence, transmitted data packets addressed to HST according to its IP address, may be dropped if the handover processing time is too long. HSTs suffer from the weak achievement of wireless connection services as trains run at high speed. Hence, this poor performance is reflected in the provided Internet services to the passengers in the compartments of those trains. The HSTs have witnessed a fast evolution in traveling at high speed up to 98 m/s and greater [129]. The provided Internet services to HSTs are far from adequate due to too much repeated handover during their rapid movement. This frequent handover means the Internet services and connectivity have become critical. A new mechanism for routing and delivering data must be implemented to ensure high data rates and continuous Internet connectivity for end-users [130]. The contribution concerns the fixed or wired part of wireless network systems (i.e., neglecting RF or wireless matter). In other words, The stationary infrastructure devices of the wireless network (routers and switches) are investigated. The proposal for designing a virtualized domain as a part of a wireless network, a domain that consists of SDN controllers, enabled OpenFlow switches and wireless access points. The main contributions in this chapter are as follows:

1. Using an SDN technique to administer the handling of packet flows during HST mobility. In SDN networks, the most important feature is to relieve the overhead control messages or signals that direct the flows of packets.

2. Drawing on the excellent feature of the SDN of its the capability of dealing with per packet or per flow independently from the IP address to destine data packets, the proposed network puts forward a technique for picking specific fields from the IP address fields to be fragmented and consolidated with other fields by utilizing SDNc to steer packet flows.

3. Utilizing a triggering signal to direct packet streams proactively. The SDNc uses the information that is included in the triggering signal to initiate redirection processing to change the packet flow paths from APc to APn. This proactive processing leads to reduced packet loss and decreased handling delay time, whilst keeping the continuity of connection during HST mobility.

5.1.1 Related Works

The world has witnessed a significant development in the maximum moving speed of HSTs. With this development, several issues have faced HSTs like a train running safety, high rate of handover repetition, and maintaining continuous Internet services. Our study concerns handling and routing packets in an SDN network. Specially, managing and controlling of data packets that have been proactively destined by an SDNc through pre-exchange controlling messages are the objectives.

The authors in [131] explored the evolution of wireless Heterogeneous Networks towards mobile 5G networks for high-speed moving vehicles. They used narrowband RF channels for the control data, while for the user traffic wideband high-frequency wireless channels were deployed. However, the drawback of this research was the utilization of different RF technologies, because forwarding of packets would need more processing, , thus resulting in long handover latency. In [132], the authors presented various scenarios for high-density transportation moving targets at high or low speed over heterogeneous networks and dense cells. They aimed to estimate the performance of a proposed vehicular networks architecture based on a multi-access router. The shortcoming of this study is dependence on the multi-access router, consisting of several antennae that could physically connect many types of wireless network technologies. Hence, a long handover process problem remained. Based on the logical design for network slicing and mobility management amongst many kinds of access points, a study was presented by [81]. This work was founded on separation of the logical network into various layers based on an edge cloud and core cloud. [133] presented a study of the causes of random delay of control services of high-speed railways and their influence on speed profile and the path of HST. These authors formulated the movement of trains as a part of the Markov approach to analyses fading channels for HSTs.

Several protocols have been proposed and implemented to resolve the handover latency and data loss dilemmas. Some them have used a reactive approach, whilst others have adopted a proactive one in relation to handling data packets. MIPv6, PMIPv6, and Hierarchical MIPv6 (HMIPv6) were designed as reactive protocols to address mobility management problems during handover within sub-networks [125, 134, 135]. These protocols have been used by communication networks to optimize the handover procedure, decrease overhead signaling, and realize layer 3 handover. Whilst, enhancements have been made in handover procedures, they still suffer from high rates of packet loss and long delay times [136]. Those protocols did not use layer 2 information to predict the direction of a moving target. As expansion and improved versions of the MIPv6 and PMIPv6 are Fast MIPv6 (FMIPv6) and Fast PMIPv6 (FPMIPv6) respectively [127]. A proactive scheme was used by these protocols, which were created principally to cope with the handover delay time and data packet loss problems. By employing a predictive scheme, the network starts layer 2 handover over using the Received Signal Strength (RSS) of the moving target [136]. All of these protocols still have considerable problems in terms of long handover latency and high rates of data loss. Moreover, They were designed to manage the mobility of low speed moving targets up to a maximum 70 km/hr [136]. Designing of 5G-Crosshaul based on SDN and NFV have been presented by [137], with this study considering the principal parameters of application elements and their interactions with the control layer. In addition, the mmWave communication network and HST scenarios were investigated by the authors. The authors in [138] put an expectation maximization algorithm depending on the historical information basis, as well they provided estimator of a blind channel model for the uplink of an RF transmission scheme on HST.

5.2 Preliminary

In this Section, a brief outline of some technologies that can be used in the proposed network to facilitate the understanding of the main idea, including object speed calculations, millimeter wave AP, massive MIMO antennae, wireless Gigabit AP, identifiers of network and host as well as IP address and mobility. In addition, a review of important related studies is provided.

5.2.1 Target Speed and Wireless Registration

High-speed moving targets (trains or cars) suffer from interrupted wireless services when they cross from the attached Access Point (AP) to a new one. That is, this happens due to breaking of the connection from APc and connecting to the APn, i.e., the wireless re-registration process. Suppose that the speed of a moving target is (100-350) km/h, the delay time of the layer 2 wireless registration Control Plane (CP) is (50-250) msec [139] and the separation distance between any two adjacent APs is (200-1000) m.

With a simple calculation via Equation 5.1, we can estimate the delay time that is required by a moving target to move from one AP to another as flows, such that:

$$t = d/v \tag{5.1}$$

where, t is the time required to cross the distance d with target speed v.

- Separation distance between APs is 200 m:
  350 km/h = 97.222 m/sec
  So, 200 m / 97.222 m/sec = 2.057 sec.

- Separation distance between APs is 1000 m:
  350 km/h = 97.222 m/sec
  So, 1000 m / 97.222 m/sec = 10.286 sec

It is clear that layer 2 wireless registration time delay of CP is much smaller than that required by a moving target to move between the adjacent APs. Consequently, there is enough time to complete RF registration when the moving target travels between any pair of adjacent APs. In other words, there is enough time to handle data packets between any pair of adjacent APs.

5.2.2 Millimeter Wave Access Point

The most important feature of a Millimeter Wave (mmW) is its frequency bands ability to carry multigigabit throughput data rates at a range from one meter to a few thousand meters. [140, 21]. The frequency bands (30-300) GHz of mmW can give more than 200 folds than the usage of current wireless bandwidth. This wide bandwidth has encouraged the wireless industry to utilize mmW APs for use in outdoor and indoor small cells ranging from a few meters to a few kilometers [141, 24]. By using mmW, a multi-Gbps rate can be delivered to the users by the beamforming technique which overcoming short and loss of channel propagation of the mmW [142, 143].

The frequency bands (30-300) GHz of mmW can give more than 200 folds than the usage of current wireless bandwidth. That wide bandwidth has urged the wireless industry to utilise the mmW APs which are the candidate to be used in outdoor and indoor small cells ranging from a few meters to a few kilometers [141], [24]. By using mmW multi-Gbps rate can be delivered to the users by beamforming technique supporting to overcome short and loss channel propagation of the mmW. [142, 143].

5.2.3 Massive MIMO Antenna

Massive Multiple Input Multiple Output (MIMO) antennae schemes can be employed to enhance the data rate, improve fidelity, boost radiated energy-efficiency, reduce of latency on the wireless interface, and simplify the multiple- access of the Media Access Control (MAC) layer. Massive MIMO impacts the performance of wireless communication systems when the sender and recipient support utilizing many antennae to receive and transmit multiple streams of data concurrently [144, 145]. By using physical layer characteristics, the massive MIMO technology can directly be combined with that of mmW to provide the throughput of Gbps traffic for wireless APs. Moreover, massive MIMO with mmW can improve link reliability [146].

5.2.4 Wireless Gigabit Attachment Point

One of the most important characteristics of the 5G network is its ability to perform high data rate ranging (1 to 7 Gbps). This means that the last distribution point of a wireless LAN (WLAN) should be able to provide a capacity of data rate of more than one Gbps to users. Wireless Gigabit (WiGig) is the standard of IEEE 802.11ad works as indoor mmW AP at 60 GHz [147], being able to achieve the multigigabit transmission rate. The beamforming technique has been found to overcome the issues of non-line-of-sight, whilst improve energy efficiency [148]. WiGig access points are suitable for being used inside the compartments of the HST.

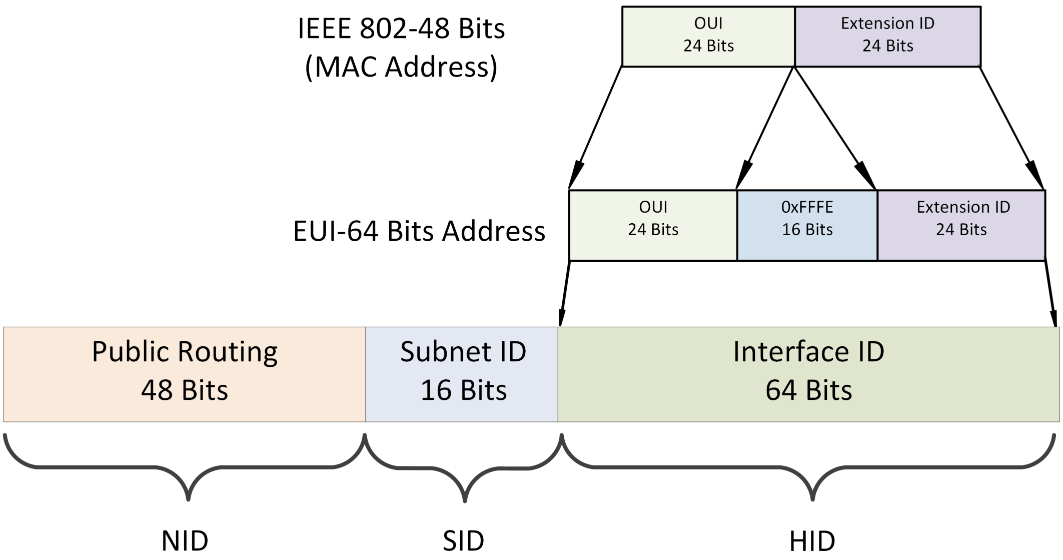


Fig. 5.1 IPv6 Structure Consists of the Extend Form of MAC Address

5.2.5 IP Address and Mobility

We have focused on IPv6, because of its capability to provide a vast number of IP addresses. Moreover, it has the significant attribute IP address auto-configuration. Also, the IPv6 address is a combination of physical ID and logical ID, which represent host and network interface IDs respectively. Fig. 5.1 illustrates the IPv6 structure, which shows how the device can identify itself through its IPv6 address. In other words, IPv6 is not utilized by the communication networks for routing purposes only, for it can also be worked as an indicator of the location of a mobile device in the network. Fig. 5.1 shows the three parts of the identifier elements of the IPv6. [149, 150]:

- Network Identifier(NID), which represents the public topology (public routing) and consists of 48 bits. Moreover, it can be considered as the parent of the SID;
- Subnetwork Identifier(SID), which pertains to the site topology (local routing) and consists of 16 bits. As above alluded to, the SID is viewed as the child of the NID which can involve a vast number of SIDs. The NID and SID together represent the global routing prefix of an IPv6. Also, they can be used as an indicator of the location;
- Host Identifier (HID), which represents the interface attachment port of a device that can connect to a network through it, being 64 bits long. The combination of the NID, SID, and

HID create a single IPv6 address that is used by a device to be an End-to-End identifier for it. Furthermore, HID can work as an indicator of the mobile device's position.

The network operator who provides services to a mobile device should have information about geographic locations of all the subnetworks that belong to his/her network. That position of a mobile device can be determined by its IP address, due to the previous location knowledge of the subnet that serves the device [14].

5.3 Proposed System Description

The structure of the proposed system as shown in Fig. 5.3 consists of three principal parts. The first represents the main domain (operator network), the second is the sub-domains (subnetworks or subnets), and part three pertains to the train (its AP and triggering signal). The following subsections explain these parts.

5.3.1 Main Domain

The main domain comprises the Software-Defined Networking controller (SDNc), Gateway (GW), Aggregator OpenFlow Switch (AOFS) and other OpenFlow Switches (OFSs) in each domain. The SDNc regulates and supervises the AOFS and other OFSs. Based on the saved information in the SDNc lookup table about the links that connect all devices associated with the system. The SDNc dictates the AOFS and OFSs to govern traffic flow from/to the main domain and sub-domains. According to this information, the SDNc creates and maintains its lookup table and makes the rules and actions, which are sent to the AOFS and other OFSs as flow tables. The SDNc is considered as the brain of the main domain due to it determining the packet routes. It sends the decisions and actions regarding these to the AOFS and OFSs to forward the data to a specific interface that links the AP to deliver the data to the destination target. That is, the decisions are taken based on the lookup tables and saved in SDNc. The lookup table of the SDNc consists of the main domain prefix (network operator prefix NID), sub-domain identifier (SID), virtual local area networks (VLANs) which serve as APs, and the train identifier which is equivalent to HID. Table 5.1 illustrates the entries of the lookup table and

Table 5.1 The SDNc Entries Lookup Table.

Main Domain Prefix (NID)	**Sub-domain (SID)**	**VLAN**	**Train ID (HID)**
2001:0DB8:ACAD/48	:0001/16	AP1	$T_{11}, T_{12}, ..., T_{1n}$
		AP2	$T_{21}, T_{22}, ..., T_{2n}$
		AP3	$T_{31}, T_{32}, ..., T_{3n}$
	:0002/16	AP4	$T_{41}, T_{42}, ..., T_{4n}$
		AP5	$T_{51}, T_{52}, ..., T_{5n}$
		AP6	$T_{61}, T_{62}, ..., T_{6n}$

the field entries that should be processed to make decisions and rules. In our proposal, the VLAN ID represents the local network topology of an AP that serves trains under its coverage area.

The AOFS receives the data from the GW to forward that data to its destination based on the SDNc rules and actions.The AOFS's responsibility is that of directing packet flow among the sub-domains as the train traverses from one sub-domain to another. That is, when the train moves from one domain to another, the AOFS changes the path of packet flow from the previous domain to the new one. However, when the train moves from one AP to another within one sub-domain, the AOFS keeps the path of packet flow unchanged. Fig. 5.2 illustrates the entries of the flow table and the fields that the packets have to subject to it. The GW represents the gate to all IP networks that are installed outside the main domain. This is a tradition device that connects the main domain (operator network) to the Internet backbone and other IP networks. It works according to rules and policies that are applied to the traditional network devices, which means that, the GW is not subject to the rules and actions of the SDNc.

5.3.2 Sub-Domains

Every sub-domain consists of at least one OFS and several APs. For simple understanding, we provided a single OFS in each sub-domain as shown in Fig. 5.3. The OFS and AOFS communicate with each other through the forwarding layer (data plane) to send and receive the data. The OFS forwards the received packets according to rules and actions that have been made by the SDNc. However, these are not the same as those that the SDNc has made for the AOFS. The OFS steers packet flow between APs as the train moves from the APc to the APn during its passing on the

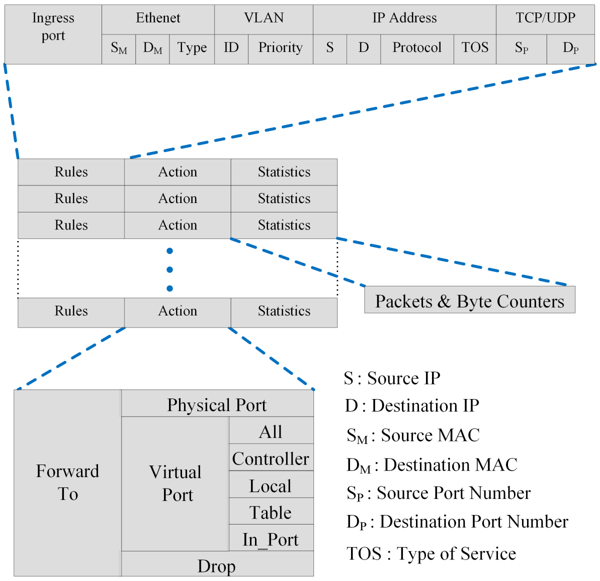


Fig. 5.2 Generated SDN OpenFlow Table

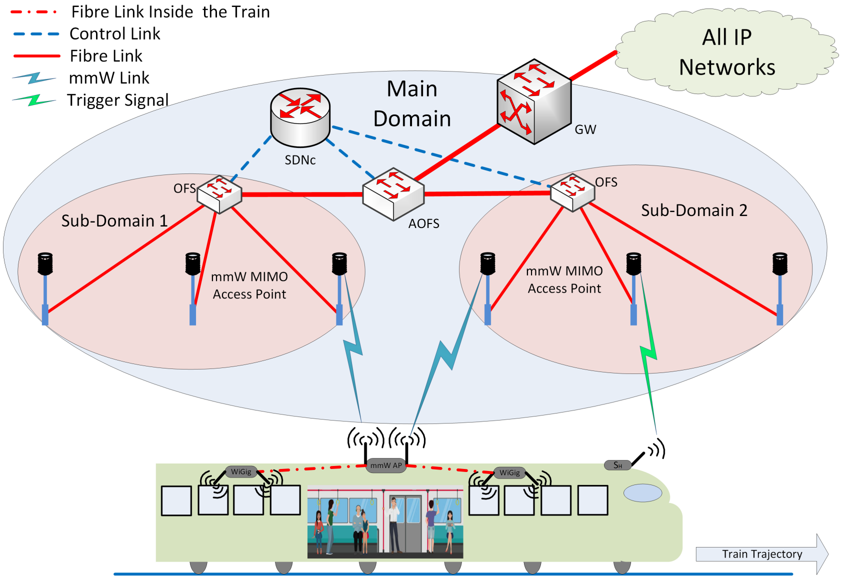


Fig. 5.3 System Architecture of High Data Rate Network for HST

trajectory. The NID is still the same for the main domain even when the train moves among the sub-domains, whereas the SID is changed by the SDNc when the train does so. APs work as VLANs, which means each AP can provide an independent virtual network that can afford many virtual IP addresses based on one real IP address. To put it simply, a single SID can provide many virtual networks, each being able to provide several IP addresses. As we mentioned above, the train AP represents HID and hence, the stateless auto-configure of the train's IPv6 address involves NID, SID, and HID. This IPv6 address indeed is used as the train identifier in a specific network. Furthermore, it can be used by the network operator as a pointer to the location of the train.

5.3.3 Trigger Signal and Train AP

The proposed system presents an innovative mechanism to manage the forwarding of packets according to an independent triggering signal that is sent by the train. This signal is received by the APn which in turn sends the triggering signal to inform the SDNc about the train location. We called this signal the handover triggering signal (S_H). The S_H is sent by an antenna positioned at the front of the train.The S_H is sent by the train in one direction to be received by the APn before the train reaches the APn so as to prepare the flows path switching of packets by the SDNc. S_H involves information such as HID, which pertains to the AP ID of the train, the destination of the train, the current AP attached to the train, and the train speed T_s. The APn forwards this information to the SDNc to switch the flow direction of the OFS only or the OFSs and the AOFS, if the train moves from one AP to another within one sub-domain or from one sub-domain to another respectively. The primary purpose of S_H is that of notifying the SDNc to dictate to the OFSs and AOFS to change the flow direction. When the SDNc receives the triggering signal, it starts to modify the flow path direction from the APc to the APn. The modifications are sent to AOFS and OFSs to achieve packet flow handover to the appropriate AP. Fig. 5.4 shows the hierarchical architecture of the proposed scheme for forwarding packets inside the main domain. Also, Fig. 5.4 points to the NID (red), SID (green), and HID (blue) parts of the IPv6 address for the train. Receiving S_H urges the APn to start layer 2 handover by notifying the SDNc to make decisions and actions to direct the flow session of the train between the APs seamlessly. This proactive procedure enables the APn to be ready to host the connection with the train while it is still associated with the current AP. The handover begins based on changing the quality parameters of the links. We rely on the received signal strength indicator (RSSI) parameter in our proposal, which can be used by the AP to predict the distance between it and the train. So the handover procedure is started by the AP to deliver the data to the train.

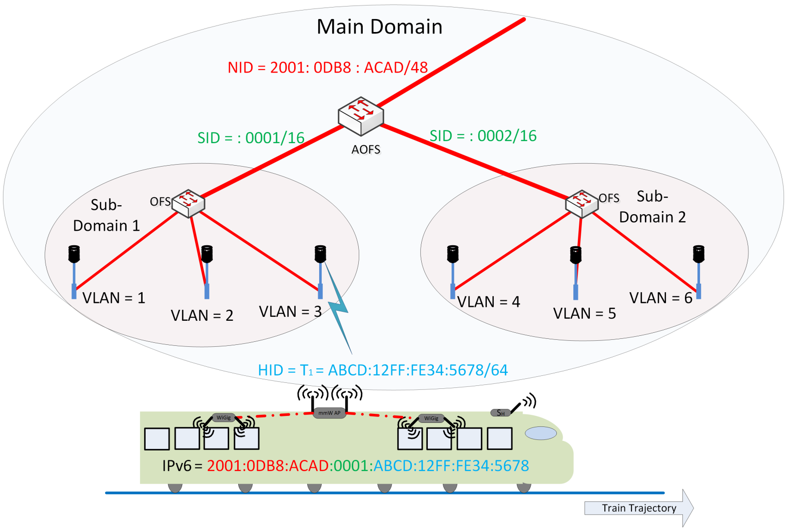
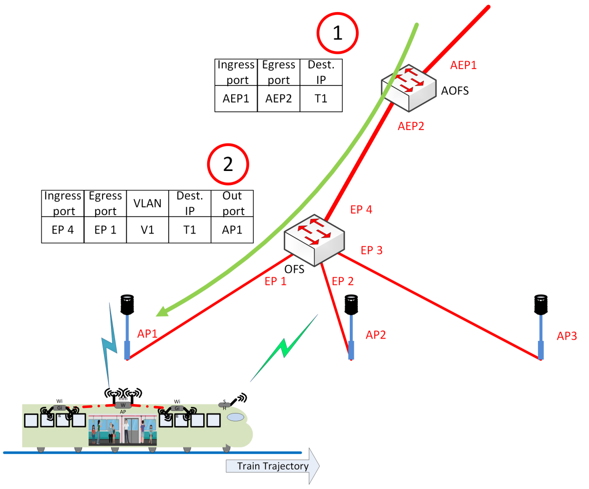


Fig. 5.4 Proposed IPv6 Hierarchy Structure for HST Network Based-SDN

5.3.4 Packet Flow Forwarding Scenario

When a mobile device changes its attachment point from one AP to another, it is required to register to the new AP at the layer 2 level. This means this registration can be made without changing the prefix (64 bits) of the IPv6 address. This process decreases the handover latency and helps to keep on-going connection without any lapse as the train moves between APs. This scenario has proposed as the train traverses between APs within one sub-domain. Fig. 5.5 represents forwarding of packet flow between the train and the APc, which is controlled by the OFS flow table created by the SDNc. This means that the AOFS flow table entries are not modified when the train transits between the APs of one sub-domain. Fig. 5.6 shows the path switching of the packet flow when the train moves from the APc to the APn in the same sub-domain. We can recognize from Figs. 5.5 and 5.6 that the responsibility of OFS is to manage this flow within one sub-domain, according to its flow table made by the SDNc. At the same time, S_H is received by the APn to provide information to commence a layer 2 handover between it and the train's AP.

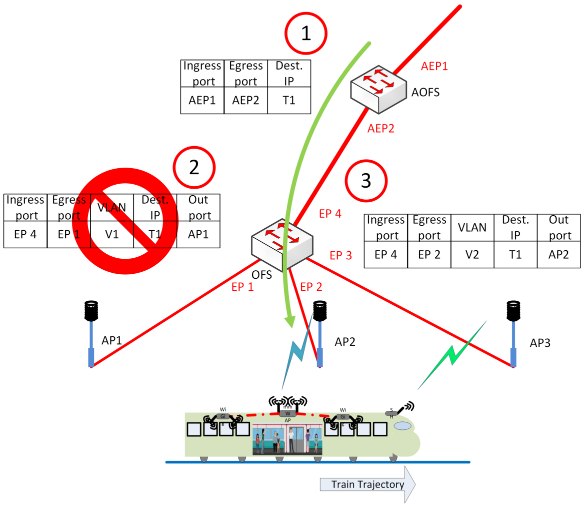
Fig. 5.7 shows the last AP of a sub-domain that is providing services to the train as the APn of another sub-domain receives the S_H. In the meantime, the flow table entries of the OFS and AOFS have not changed until the SDNc decides the right time to send modifications to the AOFS and the OFSs of the new sub-domain. Fig. 5.8 explains the events of packets handover between adjacent sub-domains. After the SDNc finishes processing the incoming information from the APn as a result



Ingress port	Egress port	Dest. IP
AEP1	AEP2	T1

Ingress port	Egress port	VLAN	Dest. IP	Out port
EP 4	EP 1	V1	T1	AP1

Fig. 5.5 Packet Forwarding Based on the Flow Table Within One Sub-domain



Ingress port	Egress port	Dest. IP
AEP1	AEP2	T1

Ingress port	Egress port	VLAN	Dest. IP	Out port
EP 4	EP 1	V1	T1	AP1

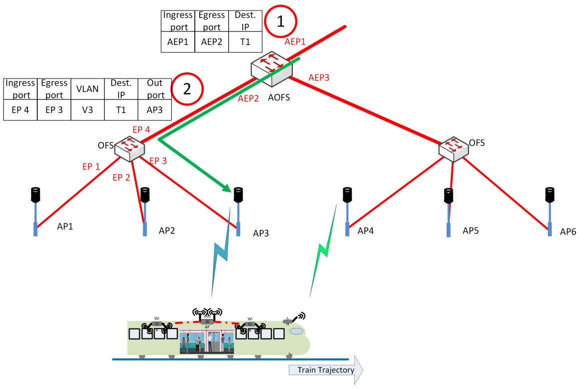
Ingress port	Egress port	VLAN	Dest. IP	Out port
EP 4	EP 2	V2	T1	AP2

Fig. 5.6 Packet Forwarding Based on the Modified Flow Table Within One Sub-domain

of the train's movement, the SDNc sends amended flow tables to the AOFS and OFS to handle the packet flow.

5.4 Distance Estimation and RSSI

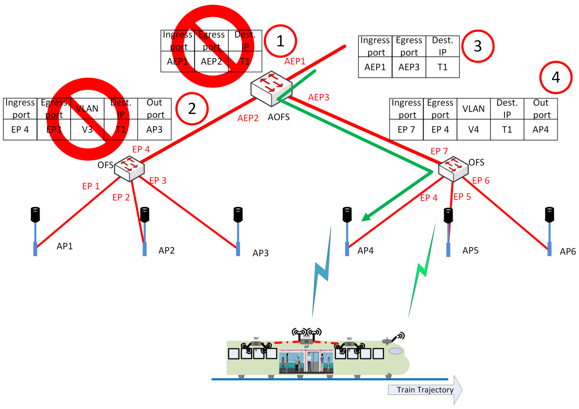
The SDNc has a holistic view of the SDN network topology and knows about the global positioning system (GPS) coordinates of the APs, the train speed and the train destination, HID and QoS provision of all the SDN network's APs. The SDNc can estimate the suitable time to be associated with the coverage range of the APn through measurements that are executed by the SDNc. These



Ingress port	Egress port	Dest. IP
AEP1	AEP2	T1

Ingress port	Egress port	VLAN	Dest. IP	Out port
EP 4	EP 3	V3	T1	AP3

Fig. 5.7 Packet Forwarding Based on the Flow Table Between Adjacent Sub-domains



Ingress port	Egress port	Dest. IP
AEP1	AEP2	T1

Ingress port	Egress port	VLAN	Dest. IP	Out port
EP 4	EP 3	V3	T1	AP3

Ingress port	Egress port	Dest. IP
AEP1	AEP3	T1

Ingress port	Egress port	VLAN	Dest. IP	Out port
EP 7	EP 4	V4	T1	AP4

Fig. 5.8 Packet Forwarding Based on the Modified Flow Table Between Adjacent Sub-domains

measurements are based on the information that is sent by S_H as well as. that programmed and saved in the applications layer and the control layer of the SDN network.

Fig. 5.9 illustrates the essential time sequence of the proposed procedure. The begins when the APn receives S_H and then forwards this information to the OFS which forwards to the SDNc to change the path of that session based on modifying the OFS flow table. This procedure occurs when the train moves between two APs within one sub-domain. While, when the train crosses from one sub-domain to another, the SDNc dictates its rules and actions to the AOFS and OFS to alter the session data flow path based on the modified flow tables on the AOFS and OFS. The SDNc measures the distance between the train and the APn by using the GPS coordinates of the two. The two-dimensional space distance separating two positions can be calculated by the famous Pythagoras theorem and can be expressed by Equation 5.2.

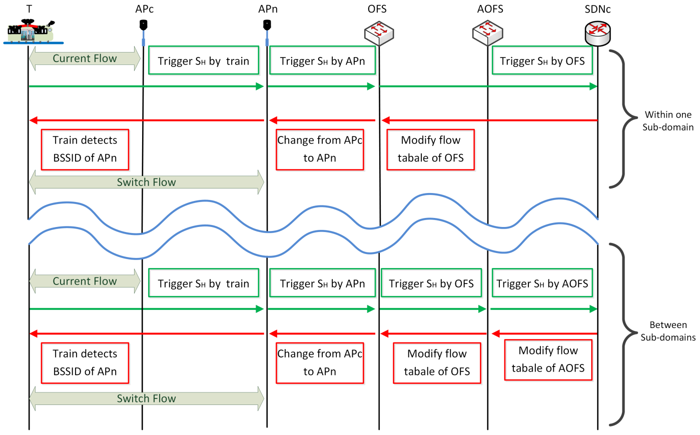


Fig. 5.9 Association with the APn: With and Without the Support of S_H

$$D_{T_i}^{APn} = \sqrt{(x_1 - x_{Ti})^2 + (y_1 - y_{Ti})^2} \tag{5.2}$$

where, $D_{T_i}^{APn}$ is the distance between the train i and the APn, x_1 and y_1 are the coordinates of the APn, whilst x_{Ti} and y_{Ti} are the GPS coordinates of the train i.

Measuring the RSSI does not refer directly to determining the position of the train. However, it can be used to form an equation in terms of changing the distance between the train and the APn. Suppose the train sent S_H with power $P_{ReD_{T_i}^{APn}}$ at the distance $D_{T_i}^{APn}$, $P_{Re_{D_0}}$ is the received signal strength at a specific known distance D_0, and the wireless wave propagation in free path loss γ equals 2 [151]. We can get Equation 5.3 as follows.

$$P_{Re_{D_{T_i}^{APn}}} = P_{Re_{D_0}} - 10 \times \gamma \times log\left(\frac{D_{T_i}^{APn}}{D_0}\right) \tag{5.3}$$

The conversion from the (mW) to the (dBm), or from (dBm) to (mW) can be done by Equations 5.4 or 5.5 respectively,

$$P_{Re_{D_{T_i}^{APn}}}(dBm) = 10 \times log[P_{Re_{D_{T_i}^{APn}}}(mW)] \tag{5.4}$$

$$P_{Re_{D_{T_i}^{APn}}}(mW) = 10^{\frac{P_{Re_{D_{T_i}^{APn}}}(dBm)}{10}} \tag{5.5}$$

Equation 5.3 can be written as follows, when γ equals 2.

$$\frac{P_{Re_{D_{T_i}^{APn}}}}{P_{Re_{D_0}}} = \left(\frac{D_{T_i}^{APn}}{D_0} \right)^2 \tag{5.6}$$

Thus, the distance between the train and the APn can be found by the RSS (S_H) at the distance $D_{T_i}^{APn}$ and the RSS ($S_{H_{D_0}}$) at the reference distance D_0,

$$D_{T_i}^{APn} = D_0 \times \sqrt{\frac{S_H}{S_{H_{D_0}}}} \tag{5.7}$$

Equation 5.7 can be used to determine the distance between the train and the APn ($D_{T_i}^{APn}$). Also, it is utilized to discover the position of the train for making the handover decision. The $D_{T_i}^{APn}$ calculated by Equation 5.7 is based on the geographic points of the X as well as Y axes, and the RSS of the S_H received by the APn with the help of GPS. All the locations of APs are known and at fixed positions in advance, being saved by the SDNc which can use this information to steer packet flows in the SDN network.

From the predicted $D_{T_i}^{nA}$ and the known T_s, the SDNc can decide the appropriate time to change the packet flow between APs. Fig. 5.10 shows the impact of utilizing S_H on link switching when the train enters a signals overlap area (it is 40 m where the lowest transmitted signals of APs can be received by the HST) of APc and APn. By using the S_H, the link switch starts at an earlier time than when not deploying it. In other words, the start of the link switch is prepared and executed by the SDNc as the APn receives S_H even the signal strength of APc is larger than APn. Using S_H improves the seamlessness handover due to the proactive preparation of layer 2 handover in SDN network. Consequently, regarding packet loss, throughput data rate, session interruption, and handover delay all these terms are enhanced.

Also, Fig. 5.10 shows the impact of using S_H on receiving data from the APn when the HST enters the overlap area. By utilizing it, the HST can receive data at almost -80 dBm, which represents the minimum value of the Transmitted Signal Strength (TSS), i.e., the transmitted power by the APn to be obtained the data from the APn by the HST. This TSS value can practically be detected and the data extracted by the HST. Moreover, from Fig. 5.10, it can be observed that the support of S_H enables the HST to receive data from APn at distance of 20 meters before the crossing point of the APc and

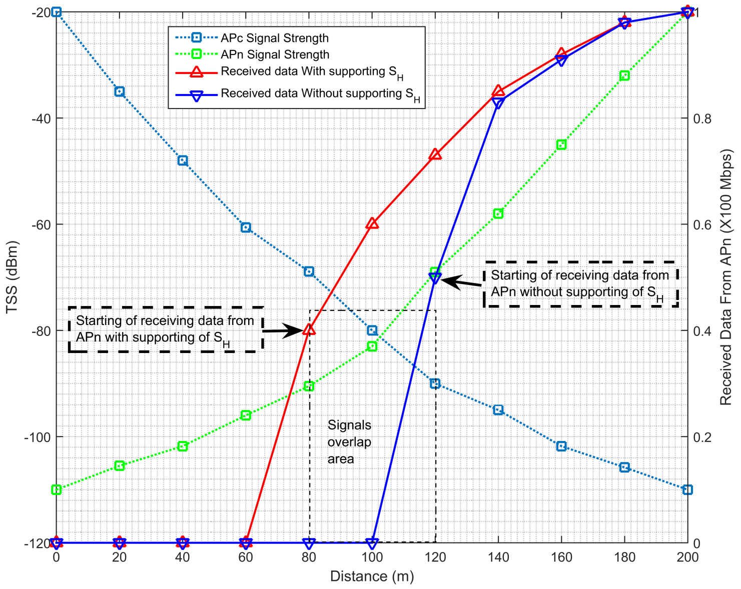


Fig. 5.10 Association with the APn: with and Without the Support of S_H

APn signals at mid-distance between them. The amount of data the HST has obtained is almost 40 Mb of the overall data of 40.8 Mb that has been injected by the simulated scenario along the overlap area. Whilst, the HST could not obtain any data from the APn when not supported by S_H until the TSS of APn hit the value of -70 dBm. This improvement in receiving the data has been achieved through proactive triggering of the SDNc to change packet flow direction from APc to the APn. Without using S_H the link switch (switch from APc to APn) only happens when the train was in the area where the TSS of APn is larger than that of APc. In this case, the SDNc receives delayed information about the train's position in preparation for changing the path of the packet flow from APc to APn.

5.5 Numerical Measurement

To determine the amount of data that exists within one meter for a specific throughput with respect to a specific T_s, Equation 5.8 can be used,

$$M_d(Mb/m) = \frac{Thr(Mb/s)}{T_s(m/s)} \tag{5.8}$$

where, M_d is the amount of data per meter regarding a certain throughput Thr of a particular bandwidth. This is with supposing the APs signals overlap area is 40 m and RF registration is 50 ms as a minimum value of layer 2 RF registration delay time. The maximum probability of packet loss within the overlap space (40 m) can be calculated by the following equation.

$$P_{loss} = \frac{DM_{overlap} \times RF_{50ms}}{1518 \times 8} \times \frac{1}{DR_{overlap}} \tag{5.9}$$

where, $DM_{overlap}$ is the amount data that can be transmitted by APs at a certain T_s and bandwidth, RF_{50ms} is the time delay of RF registration and $DR_{overlap}$ is the average received data when the train is within the signals overlap area. Table 5.2 shows the calculated numbers based on our proposal. The amount of data per meter that should be transmitted by APc or by APn within the overlap area has been estimated by Equation 5.9, according to different bandwidths and different train speeds (as shown in Table 5.2). These calculations have been used to predict the packet loss and the average of received data.

Table 5.2 The Results of Measurement Calculations

T_s	Bandwidth			Data amount within 40 m overlap area	Distance crossed within 50 ms	Data amount within RF registration delay 50
	100 Mbps	1 Gbps	10 Gbps			
98 m/s	1.02 Mb/m	10.2 Mb/m	102 Mb/m	40.8 Mb for (100 Mbps)	4.9 m	4.998 Mb for (100 Mbps)
				408 Mb for (1 Gbps)		49.98 Mb for (1 Gbps)
				4080 Mb for (10 Gbps)		499.8 Mb for (10 Gbps)
56 m/s	1.786 Mb/m	17.86 Mb/m	178.6 Mb/m	71.44 Mb for (100 Mbps)	2.8 m	5.0008 Mb for (100 Mbps)
				714.4 Mb for (1 Gbps)		50.008 Mb for (1 Gbps)
				7144 Mb for (10 Gbps)		500.08 Mb for (10 Gbps)
28 m/s	3.571 Mb/m	35.71 Mb/m	357.1 Mb/m	142.84 Mb for (100 Mbps)	1.4	4.9994 Mb for (100 Mbps)
				1428.4 Mb for (1 Gbps)		49.994 Mb for (1 Gbps)
				14284 Mb for (10 Gbps)		499.94 Mb for (10 Gbps)

Fig. 5.11 illustrates the percentage of the probable number of lost packets within the overlap distance. The figure shows that the rate of the lost packets increases due to a rise in the capacity of the channel bandwidth and with increasing the speed of the train. The increase in the percentage of lost packets is as a result of the increasing in the data amount (Mb) that existed in the distance unit (m). This means the APc or APn transmit data rate per time unit per meter (Mb/s/m) depends on channel bandwidth capacity, hence, every meter crossed by the HST is being not connected to APc either APn will increase the number of packet loss. As well as, the packet loss increases with an increase in the HST's speed, i.e., the traversed meters by the HST for one second will be higher when it moves a

quickly than it runs slowly. That is, the HST crosses relatively long distance without it is neither associated with APc nor APn. By way of explanation, the number of lost packets depends on the layer 2 RF registration time that the train needs to connect with APn. That is, a long RF registration time leads to a large number of packets being lost. Furthermore, when the train moves fast, it will move a longer distance than when it slower, but the same amount of time is required to complete the RF registration (50 ms to associate with APn).

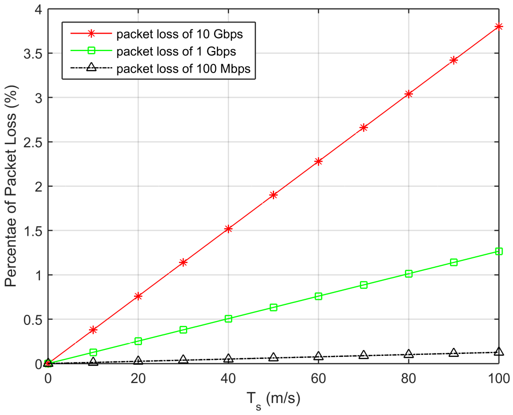


Fig. 5.11 Lost Packets for Different Throughput vs. T_s

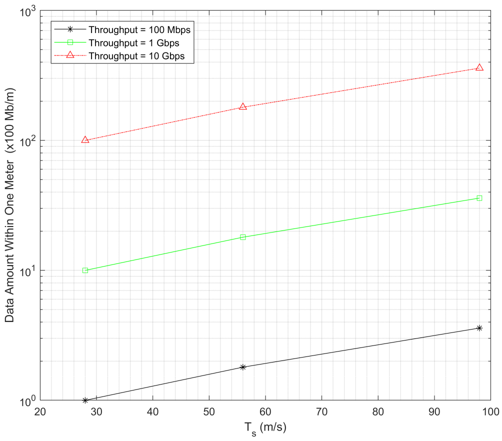


Fig. 5.12 Data Amount per Meter Against Different Values of T_s

Fig. 5.12 shows the data amount per meter regarding different values of a channel throughput versus different train speeds that related to 50 ms. The data amount per meter increases exponentially with increasing channel throughput. In other words, the number of bits in each meter jumps to a high value due to the capacity of the transmission channel (100 Mbps, 1 Gbps, and 10 Gbps). On the other hand, an increase in the T_s (as proposed 28, 56, and 98) m/s leads to an increase in the number of meters that are traversed by the HST within 50 ms.

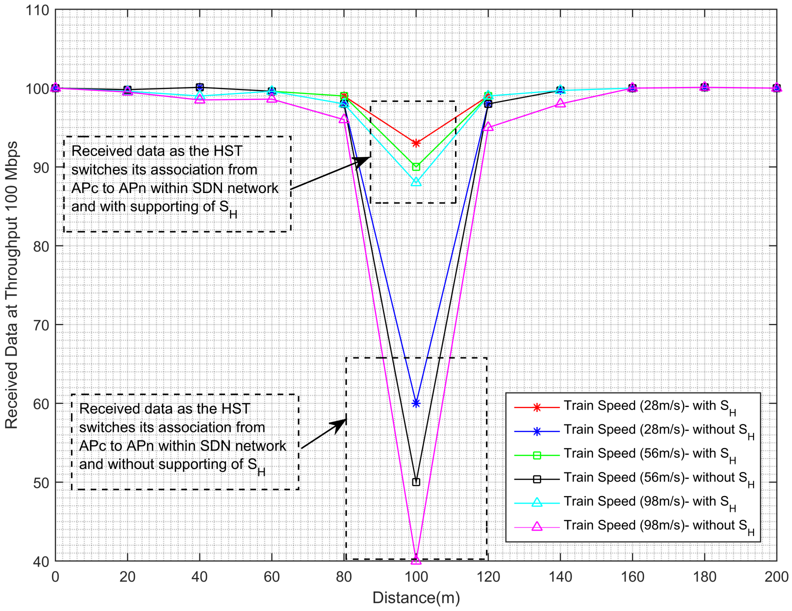


Fig. 5.13 Average Received Data During HST Crosses Signals Overlap Area

Fig. 5.13 shows the received data against T_s during the HST traverses between the APc and the APn within the proposed SDN network. Supposing the throughput of the channel capacity is 100 Mbps as the throughput data rate. Then, the obtained results show that the received data with supporting S_H are higher than those reported without it as a proactive signal to start path modification by the SDNc. The proposed an overlap distance is 40 m (between 80 m and 120 m) for collecting the data. The line graph above represents the average data received by the HST from APc and APn together. It can be seen that the proactive scheme of layer 2 RF registration improved the SDN network performance in handling the packet flow between the APs. This improved performance appeared in the received data at the signals overlap area. The HST received 93%, 90% and 89% of the data that existed in the 40 m of overlap distance at different values of T_s of 28 m/s, 56 m/s, and 98 m/s, respectively. In contrast, the received data percentages were 61%, 52%, and 40% before using

S_H. This degradation in received data percentages is as a result of the reactive system of the SDNc to direct the packet flow between the APs without the support of S_H.

To build and manage a dynamic network in an SDN environment. The SDNc should exchange control messages (packets) with the other devices (OFS)s to handle the data flow within the network. We have considered the delay (transmission, process, and queue) in successive devices along the whole path instead of a delay in each device belonging to that path. By pinging the SDNc from any AP (host), we can notice the overall delay time of packets within the round trip time required to finish its journey. This constitutes three delay times making up the total (t_{tot}), namely: the RTT between the SDNc and the OSF (t_{OFS}), the RTT between the SDNc and the AOSF (t_{AOFS}), and the link delay (t_{link}). We represented the delay time by the following equation:

$$t_{tot} = t_{link} + t_{OFS} + t_{AOFS} + Cal \tag{5.10}$$

where, Cal represents the SDNc calibration with variable value and corresponds to the constraints of the SDNc. Fig. 5.14 shows that the estimated delay values grow with an increase in the number of hops. The measured values are obtained from the hop number multiplied by the offset value of the SDNc [152]. Moreover, it can be seen from Fig. 5.14 that the path delay is linear with increasing the number of the switches that are linked serially in the SDN network.

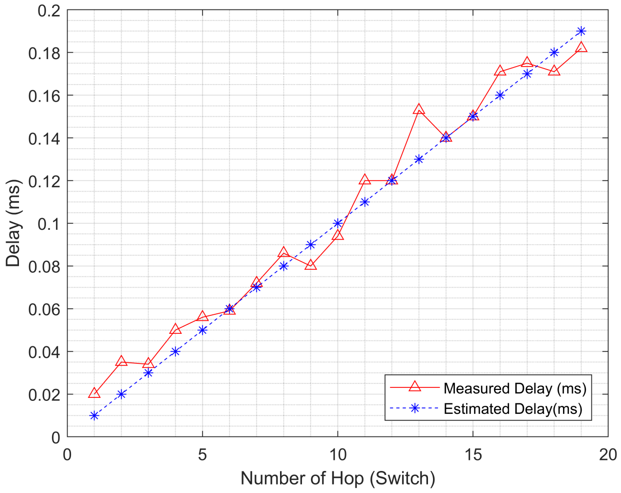


Fig. 5.14 Path Delay of the Control Packets Corresponding to the Number of Switches

5.6 Testbed Setup and Implementation

Mininet simulator, Python network programming, and MATLAB software have been used to virtually realization and implementation of the proposed SDN network with supporting physical APs devices.

1. Mininet Simulator

   Mininet simulator is a network emulator that can design and imitate a complete OpenFlow network locally on a desktop or laptop computers. Mininet simulator is a network emulator which builds virtually network devices such as hosts, links, switches, and controllers. The hosts run Linux system software. The switches support OpenFlow for profoundly adaptable manner routing and SDN. Mininet supports research, learning, development, prototyping a network, testing and performing a complete experimental network on a desktop or laptop. Mininet affords a mild and reasonable network testbed, allows complex topology testing without wiring up any physical devices, supports random system topologies, and provides an extensible Python API for experimental network production.

2. Python Network Programming

   Python is a comprehensive object-oriented programming method including a standard library that covers everything demanded to quickly create effectual network applications. Moreover, Python has a manifold of third-party libraries and combinations that expand Python to all field of network programming and topologies creation.

3. MATLAB Platform

   The MATLAB has been employed to assess the performance of the proposed SDN network. In addition, it has been used to interface and control the physical APs devices.

To simulate the connection change between the train and APs (handover), we utilised two APs which represent (APc and APn). By regulating the transmitted signal strength (TSS) of both in which the TSS of APc decreases the TSS of APn increases at the same time and the same value simultaneously. This represents and mimics the assumed movement of the HST that should move from the APc to the APn. The crossed point of the curves in Fig. 5.10 represents the threshold point of TSS of APs. At this point, the train can receive data packets from the APc and APn. This

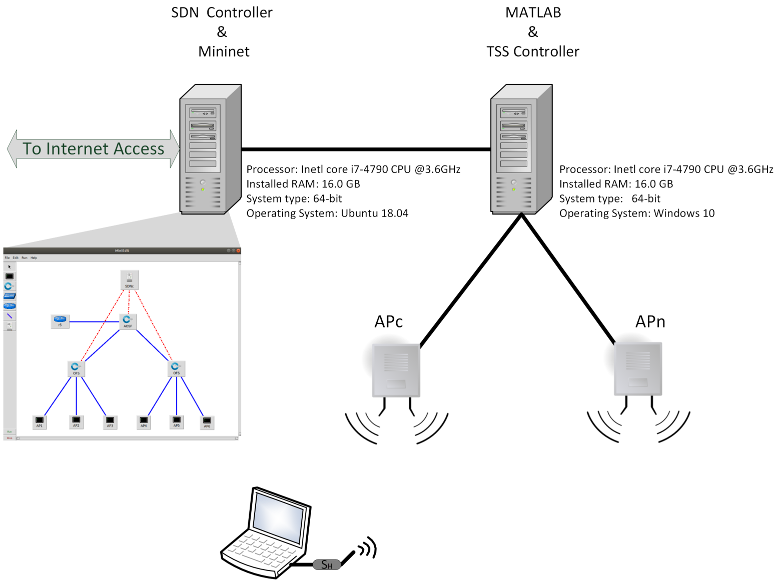


Fig. 5.15 Testbed Setup Block Diagram.

process has been simulated and coded by the MATLAB platform. 400 packets have been injected by the programming code to APs during the handover procedure. Handling the packets path direction (switching from APc to APn) based on the Mininet VM which includes the SDNc, AOFS, and OFS. We neglected the wireless registration procedure, i.e., the RF registration (Authentication, Authorization and Accounting).

Fig. 5.15 illustrates the testbed setup of the proposed network that is implemented to emulate the SDN network. The Mininet network has been configured by Python programming language. The experiment topology consists of 3 switches (2 OFS and 1 AOFS) are linked by the SDNc, and 6 hosts represent the APs are connected by OFSs. For simplicity, we connected 2 physical APs and controlled by MATLAB code as shown in Fig. 5.16.

In the first experimental scenario, we performed the measurements to see the control delay when the injected data changed its path from APc to APn without supporting of using the S_H. The second experiment scenario is achieved by the same way, but with supporting of utilising the S_H. Fig. 5.17 shows the measured delay and estimated delay for the experiment scenarios. The parameters of this experiment are listed in Table 5.4.

The performance of the proposed SDN network is improved by using the proactive scheme versus the reactive scheme for both measured and estimated controlling delay.

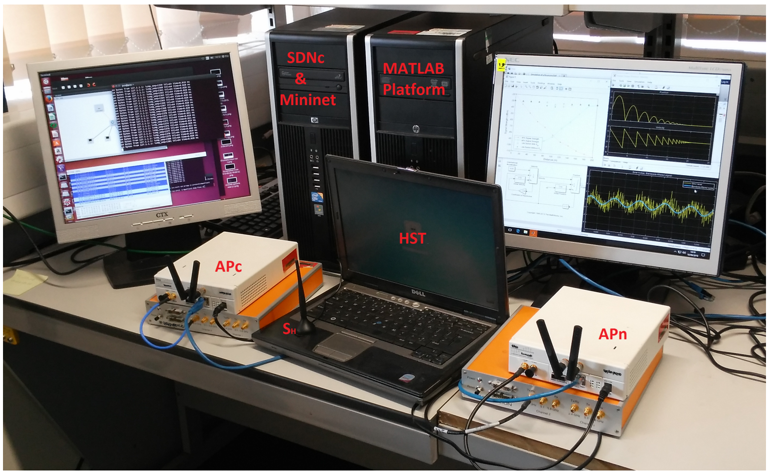


Fig. 5.16 Practical Testbed Setup of Proposed SDN Network.

The overall test period is 100 seconds. We configured and programmed the network based on the HST sends its S_H which is received by the APc, at that instant APc pings the SDNc to make the flow decision. this process is configured to happen after 20 seconds of beginning the test. To mimic the transferring of HST from APc to APn, we controlled the TSS of APc by attenuating its TSS, at the same time, the TSS of APn amplifies with respect to the time. By this way, the handover between the APs has been mimicked. The injected data starts at TSS equals (-80 dBm) to represent the proposed data that should be in the signals overlap area. The amount of data in this area is expressed as the throughput rate of proposed channel capacities (40.8 Mb to 14284 Mb for 100 Mbps and 10 Gbps respectively). The experiment setup was constructed by using two *hp* computers as workstations in this work. These workstations have the specifications as mentioned in Table 5.3. In Fig. 5.17 a comparison between the measured and estimated delay values with supporting of using S_H and without using S_H for both cases. The figure shows that the practical and predicted results scored

Table 5.3 The Hardware and Software used in Testbed

Item	**PC1**	**PC2**
Type	hp	hp
CPU Vendor	Intel(R)	Intel(R)
CPU Type	Core i7-4790 @ 3.60 GHz	Core i7-4790 @ 3.60 GHz
RAM	16 GB	16GB
Host Operating System	Windows 10	Ubuntu 14.4
Host Software	MATLAB R2017b	Python 3.7.0, Mininet 2.2.1

Table 5.4 The Testbed Setup Parameters

Parameters	Value
T_s	28, 56, 98 m/sec
Packet Size	1522 Byte
Control Messages	10 message
Overlap Distance	40 m
Delay Per Link	0.001 msec
Processing Delay	0.005 msec
OFS Modifying Delay	0.005 msec
AOFS Modifying Delay	0.005 msec
SDN Modifying Delay	0.010 msec
Layer 2 RF Registration	50 msec
Test Duration	100000 msec

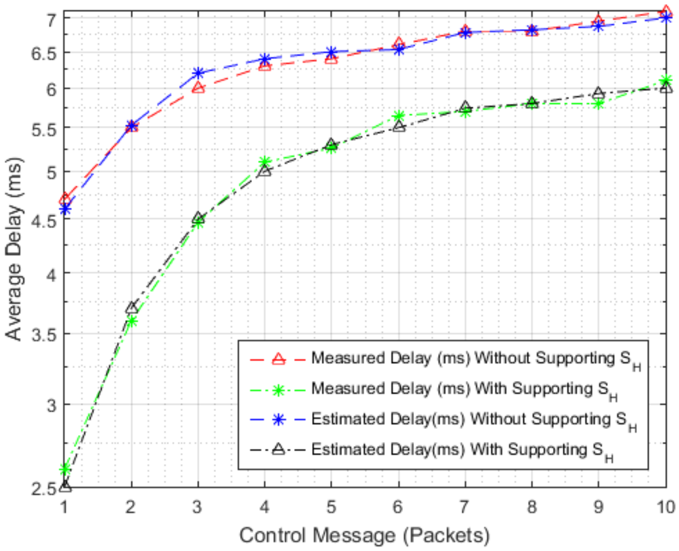


Fig. 5.17 Delay Comparison of Control Packet Exchange Between SDNc and Switches

higher delays when the S_H was not utilised as a trigger signal than the measured and estimated results when the S_H was employed. The corresponding estimation deviation varies from 0.02 ms to 0.27 ms. Also, in Fig. 5.17 can be observed that in both cases of estimated and measured results, the proactive scheme of directing the flow of packets is developed by using the S_H in SDN network. The control delay has experienced a noticeable reduction in measured values by utilising the S_H. This reduction of the delay can be observed at the launch of making the decision by the SDNc to switch the packets flow path. This decrease in the delay is due to the proactive processing of control messages that are exchanged by SDNc, AOFS, and OFSs to switch the packets stream from APc to APn. The same

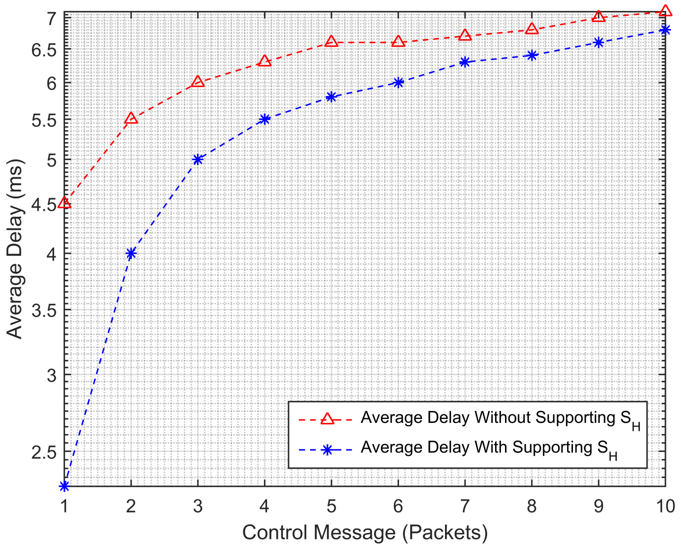


Fig. 5.18 Impact of S_H on the Average Delay of Control Packets

observation of control delay values can be seen with the predicted values. Where the control delay decreases at the beginning of taking the decision to change the packets flow path between APs.

The comparison of the effect of using S_H on the average delay values is presented in Fig. 5.18. From the results in Fig. 5.18, we find that the average delay values of experimenting values without using proactive signal is higher than that of experimenting values of average delay that are supported by the proactive signal(S_H). The average delay almost is halved by using the S_H. The SDN network can be enhanced by using a creative idea to improve the performance of the SDN network through reducing the delay of both the control plane and data plane.

5.7 Chapter Summary

This chapter has presented a novel method of handling packet flows in HST networks based on SDN and smart virtualization (SDN, NFV, and network virtualization). Packets re-directing delay, packets forwarding, and packets loss problems as the HST traverses amongst APs have been investigated. We have applied a novel proactive scheme for directing the packet flows among the APs by using trigger signal to begin layer 2 handover. The virtualization has added exceptional features to the communication networks such as cloud computing, SDN, NFV, etc. These techniques of virtualization enable the communication networks to improve its performance. Although of these techniques have been applied, the communication networks still need new creative ideas to meet the demands of the vital issues such as reducing the delay, high data rate throughput, reliability, session continuity, etc. Our proposal suggests an SDN network environment serves HST and provides a low controlling delay and high delivering of a data rate. In this chapter, a description of the proactive scheme which is employed to handle the packets flow for the HST network based-SDN. The suggested SDN network is emulated by Mininet, Python and MATLAB. Also, this chapter presents empirical and theoretical analyses of the suggested SDN network. The simulated and measured results were compared by two scenarios. The first scenario was a reactive scheme and the second one was a proactive scheme. The advantages of the proactive scheme are obtained through the reduction of controlling delay messages between the SDNc and underneath OpenFlow switches. Moreover, decrease the number of lost packets during the handover procedure in both cases of increasing in the channel bandwidth capacity, and the train speed. Based on the results, we can conclude that the obtained results indicate a necessity to find an innovative conception to enhance and support the virtualization techniques of carrying and directing the packets flow, avoiding packets loss which is extremely influenced by the speed of HST and the capacity of the transmission link bandwidth. The suggested SDN network has been emulated and performed by Mininet simulator, Python language, and MATLAB platform.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

This chapter has briefly reviewed and concluded the findings of this thesis. Also, it has made recommendations for future works that could enhance the performance of communication systems. This thesis laid the foundation to cooperating between virtualization technologies and non-standard packets forwarding approaches for different network domains. In order to reach the optimal advantages of current virtualization technologies, these technologies should be combined and supported by innovative and new ideas and solutions to create advanced communication systems. Several experimental scenarios are employed to examine and evaluate the proposed networks performance in terms of End-to-End delay time, processing delays, queueing delay time, and packet loss ratio. In these experimental scenarios, the proposed system networks are emulated, programmed, and simulated by Mininet, Python, and Matlab platforms.

In spite of all clever ideas and advanced technologies that have been deployed such as SDN, NFV, NS, and NFS, there are many obstacles that prevent meeting the 5G network specifications. One of the most vital obstacles is the control delay time or to be more specific, is the processing control latency that is caused by communications network devices. The principal finding of this thesis is the chief cause of delays in communication systems is the control layer due to its responsibilities such as,

- Exchanging control messages between associated interfaces of network devices to enable data transmission between them.

- Negotiation messages which prepare and regulate a link to transfer data packets amongst network devices.
- Network management messages, these messages administrate the physical and logical infrastructure of networks according to the view of the network administrator.

These delays are dominant in wireless networks due to registration and security issues. While in wired networks, the control delay time is somewhat less due to the fixed position of connected devices of the communication systems. The next points highlight the foremost conclusions of the work achieved in this study:

1. The average delay time of End-to-End connection of the DP is far less than the average delay time of End-to-End connection of the CP. In other words, data streams need much less time to pass through a network device than the time that the network device needs to make decisions for that data.
2. The air-interface (i.e., RF) registration in wireless networks is still the longest delay time consumed.
3. The handover and packets loss issues are still at high average values and need more improvement to minimize its effects.
4. Future communication networks should implement a new scheme of controlling and handling mechanisms of data packets. This new scheme should ensure a high data rate delivery and highly reliable connection to the Internet for end-users.
5. To ensure a high data rate and stable Internet connection to end-users, a new scheme data routing, and delivery mechanisms should be implemented.

6.2 Future Works

This thesis tiles the ground to incorporate virtualization technologies and non-standard packets routing and forwarding proposals. To obtain the optimum benefits of modern virtualization technologies. Future communication networks based-SDN-NFV-Virtualization will desegregate all limitations on the designated purpose of networks. Therefore, open sources platform will uncork doorways new network topologies and infrastructures. This thesis paves the way for advanced networks, hence the future extent of this research can be developed further by:

1. Experimenting the same network topology, but in the real environment of a specific design for network topology. Practical implementation of SDN network will give a clear view of the network performance based-SDN with support of novel thoughts.

2. Planning different networks topology for SDN controller (centralized, semi-centralized, distributed) can be tested to evaluate functionality such as scalability, convergence, manipulation, and processing.

3. Handling a real-live data stream can be investigated by forwarding it through the designed network based on creative ideas. This investigation will improve the smart television services network management.

4. Using a new and not used before methods and algorithms for data traffic engineering scheme to reduce decisions process to forward users' data. By utilizing non-standard and open sources application, the processing delay could be lessened to minimal values by designing and implementing smart algorithms to deliver the data.

5. Observing the effects of load balance by utilizing diverse data mining methods with lessened anticipate times, instead of using a neural network algorithm because it requires spending the relatively long time to process.

References

[1] T. ETSI, “123 501 v15. 2.0 (2018-06) 5g,” *System Architecture for the 5G System (Release 15)*.

[2] P. Marsch, Ö. Bulakci, O. Queseth, and M. Boldi, *5G system design: architectural and functional considerations and long term research*. John Wiley & Sons, 2018.

[3] M. Faerber, “Collaborative 5g research within the eu framework of funded research,” *Towards 5G: Applications, Requirements and Candidate Technologies*, pp. 23–33, 2016.

[4] A. Johnson, *31 Days Before Your CCNA Routing & Switching Exam: A Day-By-Day Review Guide for the ICND1/CCENT (100-105), ICND2 (200-105), and CCNA (200-125) Certification Exams*. Cisco Press, 2017.

[5] R. Perlman, *Interconnections: bridges, routers, switches, and internetworking protocols*. Pearson Education India, 1999, vol. 2, ch. 1, pp. 3–8.

[6] R. Froom and E. Frahim, *Implementing Cisco IP switched networks (SWITCH) foundation learning guide:(CCNP SWITCH 300-115)*. Cisco Press, 2015.

[7] K. Ahmed and S. Hranilovic, “C-ran uplink optimization using mixed radio and fso fronthaul,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 6, pp. 603–612, 2018.

[8] D. Chitimalla, K. Kondepu, L. Valcarenghi, M. Tornatore, and B. Mukherjee, “5g fronthaul-latency and jitter studies of cpri over ethernet,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 2, pp. 172–182, 2017.

[9] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, “5g backhaul challenges and emerging research directions: A survey,” *IEEE access*, vol. 4, pp. 1743–1766, 2016.

[10] A. Johnson, *Routing protocols and concepts, CCNA exploration labs and study guide*, 2007.

[11] R. Graziani and A. Johnson, *Routing Protocols and Concepts, CCNA exploration companion guide*. Cisco Press, 2007.

[12] G. L. White, J. R. Shah, and J. R. Cook, “Internet technology in 2010: The issue of ipv6 adoption in the usa,” *Journal of International Technology and Information Management*, vol. 14, no. 3, p. 5, 2005.

[13] R. Molenaar, *how to master ccna*. tshiamo sigwele, 2013, pp. 379–425.

[14] J. Davies, *Understanding ipv6*. Pearson Education, 2012, ch. 2–3, pp. 17–86.

[15] J. Korhonen, T. Savolainen, and J. Soininen, *Deploying IPv6 in 3GPP networks: evolving mobile broadband from 2G to LTE and beyond*. John Wiley & Sons, 2013.

[16] P. Sharma, “Evolution of mobile wireless communication networks-1g to 5g as well as future prospective of next generation communication network,” *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 8, pp. 47–53, 2013.

[17] V.-G. Nguyen, A. Brunstrom, K.-J. Grinnemo, and J. Taheri, “Sdn/nfv-based mobile packet core network architectures: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1567–1602, 2017.

[18] Y. B. Lin and A. C. Pang, *Wireless and mobile All-IP networks.* John Wiley & Sons, 2005.

[19] H.-H. Cho, C.-F. Lai, T. K. Shih, and H.-C. Chao, “Integration of sdr and sdn for 5g,” *IEEE Access*, vol. 2, pp. 1196–1204, 2014.

[20] S. Abolfazli, Z. Sanaei, S. Y. Wong, A. Tabassi, and S. Rosen, “Throughput measurement in 4g wireless data networks: Performance evaluation and validation,” in *Computer Applications & Industrial Electronics (ISCAIE), 2015 IEEE Symposium on.* IEEE, 2015, pp. 27–32.

[21] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5g cellular: It will work!” *IEEE access*, vol. 1, pp. 335–349, 2013.

[22] A. Khan, D. Jurca, K. Kozu, W. Kellerer, and M. Yabusaki, “The reconfigurable mobile network,” in *Communications Workshops (ICC), 2011 IEEE International Conference on.* IEEE, 2011, pp. 1–5.

[23] M. Series, “Imt vision–framework and overall objectives of the future development of imt for 2020 and beyond,” *Recommendation ITU*, pp. 2083–0, 2015.

[24] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[25] V. H. Muntean and M. Otesteanu, “Wimax versus lte-an overview of technical aspects for next generation networks technologies,” in *Electronics and Telecommunications (ISETC), 2010 9th International Symposium on.* IEEE, 2010, pp. 225–228.

[26] H. Holma and A. Toskala, *WCDMA for UMTS: Radio access for third generation mobile communications.* john Wiley & sons, 2005.

[27] C. Cox, “An introduction to lte, lte-advanced, sae and 4g mobile communications,” *Editorial WILEY*, 2012.

[28] A. Mansour and M. Alnas, “Fast mobile ipv6 handover using link and location information,” *International Journal of Computer Science and Security (IJCSS)*, vol. 9, no. 3, p. 174, 2015.

[29] C. E. Perkins, “Mobile ip,” *IEEE communications Magazine*, vol. 35, no. 5, pp. 84–99, 1997.

[30] H. Soliman, C. Castelluccia, K. Elmalki, and L. Bellier, “Hierarchical mobile ipv6 (hmipv6) mobility management,” Tech. Rep., 2008.

[31] M. Blanchet, *Migrating to IPv6: a practical guide to implementing IPv6 in mobile and fixed networks.* John Wiley and Sons, 2009.

[32] I. A. Alimi, A. L. Teixeira, and P. P. Monteiro, “Toward an efficient c-ran optical fronthaul for the future networks: A tutorial on technologies, requirements, challenges, and solutions,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 708–769, 2017.

[33] F. Tian, P. Zhang, and Z. Yan, "A survey on c-ran security," *IEEE Access*, vol. 5, pp. 13 372–13 386, 2017.

[34] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 11, pp. 5275–5287, 2015.

[35] T. G. Rodrigues, K. Suto, H. Nishiyama, N. Kato, and K. Temma, "Cloudlets activation scheme for scalable mobile edge computing with transmission power control and virtual machine migration," *IEEE Transactions on Computers*, vol. 67, no. 9, pp. 1287–1300, 2018.

[36] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[37] A. Wang, M. Iyer, R. Dutta, G. N. Rouskas, and I. Baldine, "Network virtualization: Technologies, perspectives, and frontiers," *Journal of Lightwave Technology*, vol. 31, no. 4, pp. 523–537, 2013.

[38] Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel, "Lte wireless virtualization and spectrum management," in *Wireless and Mobile Networking Conference (WMNC), 2010 Third Joint IFIP*. IEEE, 2010, pp. 1–6.

[39] K. Gray and T. D. Nadeau, *Network function virtualization*. Morgan Kaufmann, 2016.

[40] A. Blenk, A. Basta, M. Reisslein, and W. Kellerer, "Survey on network virtualization hypervisors for software defined networking," *arXiv preprint arXiv:1506.07275*, 2015.

[41] Z. Zhang, Z. Li, K. Wu, D. Li, H. Li, Y. Peng, and X. Lu, "Vmthunder: fast provisioning of large-scale virtual machine clusters," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3328–3338, 2014.

[42] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[43] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "Nvs: A substrate for virtualizing wireless resources in cellular networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 5, pp. 1333–1346, 2012.

[44] C. Colman-Meixner, C. Develder, M. Tornatore, and B. Mukherjee, "A survey on resiliency techniques in cloud computing infrastructures and applications," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2244–2281, 2016.

[45] P. Sanjoy and S. Srini, "Technical document on wireless virtualization," Document GDD-06-17, GENI Project, USA, Tech. Rep., 2006.

[46] A. T. Velte, T. J. Velte, R. C. Elsenpeter, and R. C. Elsenpeter, *Cloud computing: a practical approach*. McGraw-Hill New York, 2010.

[47] Z. Cai, X. Li, and J. Gupta, "Heuristics for provisioning services to workflows in xaas clouds," *IEEE Transactions on Services Computing*, no. 1, pp. 1–1, 2016.

[48] M. E. Khalil, K. Ghani, and W. Khalil, "Onion architecture: a new approach for xaas (everything-as-a service) based virtual collaborations," in *Learning and Technology Conference (L&T), 2016 13th*. IEEE, 2016, pp. 1–7.

[49] Z. Chang, Z. Zhou, S. Zhou, T. Chen, and T. Ristaniemi, “Towards service-oriented 5g: Virtualizing the networks for everything-as-a-service,” *IEEE Access*, vol. 6, pp. 1480–1489, 2018.

[50] T. Taleb, M. Corici, C. Parada, A. Jamakovic, S. Ruffino, G. Karagiannis, and T. Magedanz, “Ease: Epc as a service to ease mobile core network deployment over cloud,” *IEEE Network*, vol. 29, no. 2, pp. 78–88, 2015.

[51] Q. F. Hassan, M. Elkhodr, and S. Shahrestani, *Networks of the Future: Architectures, Technologies, and Implementations*. Chapman and Hall/CRC, 2017.

[52] J. Doherty, *SDN and NFV simplified: a visual guide to understanding software defined networks and network function virtualization*. Addison-Wesley Professional, 2016.

[53] R. Jain and S. Paul, “Network virtualization and software defined networking for cloud computing: a survey,” *IEEE Communications Magazine*, vol. 51, no. 11, pp. 24–31, 2013.

[54] B. R. Al-Kaseem and H. S. Al-Raweshidyhamed, “Sd-nfv as an energy efficient approach for m2m networks using cloud-based 6lowpan testbed,” *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1787–1797, 2017.

[55] K. Katsalis, N. Nikaein, E. Schiller, R. Favraud, and T. I. Braun, “5g architectural design patterns,” in *Communications Workshops (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 32–37.

[56] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, “5g on the horizon: key challenges for the radio-access network,” *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 47–53, 2013.

[57] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, “Nox: towards an operating system for networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 3, pp. 105–110, 2008.

[58] P. Goransson, C. Black, and T. Culver, *Software defined networks: a comprehensive approach*. Morgan Kaufmann, 2016.

[59] N. Feamster, J. Rexford, and E. Zegura, “The road to sdn,” *Queue*, vol. 11, no. 12, p. 20, 2013.

[60] N. Bizanis and F. A. Kuipers, “Sdn and virtualization solutions for the internet of things: A survey,” *IEEE Access*, vol. 4, pp. 5591–5606, 2016.

[61] M. Alenezi, K. Almustafa, and K. A. Meerja, “Cloud based sdn and nfv architectures for iot infrastructure,” *Egyptian Informatics Journal*, 2018.

[62] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, “Network function virtualization: State-of-the-art and research challenges,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 236–262, 2016.

[63] F. A. Yaseen, N. A. Al-Khalidi, and H. S. Al-Raweshidy, “Smart virtual enb (svenb) for 5g mobile communication,” in *Fog and Mobile Edge Computing (FMEC), 2017 Second International Conference on*. IEEE, 2017, pp. 88–93.

[64] M. Karakus and A. Durresi, “A survey: Control plane scalability issues and approaches in software-defined networking (sdn),” *Computer Networks*, vol. 112, pp. 279–293, 2017.

[65] F. Olivier, G. Carlos, and N. Florent, “New security architecture for iot network,” *Procedia Computer Science*, vol. 52, pp. 1028–1033, 2015.

[66] M. Yang, Y. Li, D. Jin, L. Zeng, X. Wu, and A. V. Vasilakos, "Software-defined and virtualized future mobile and wireless networks: A survey," *Mobile Networks and Applications*, vol. 20, no. 1, pp. 4–18, 2015.

[67] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.

[68] R. Alvizu, G. Maier, N. Kukreja, A. Pattavina, R. Morro, A. Capello, and C. Cavazzoni, "Comprehensive survey on t-sdn: Software-defined networking for transport networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2232–2283, 2017.

[69] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5g with sdn/nfv: concepts, architectures and challenges," *arXiv preprint arXiv:1703.04676*, 2017.

[70] Q. Yan, F. R. Yu, Q. Gong, and J. Li, "Software-defined networking (sdn) and distributed denial of service (ddos) attacks in cloud computing environments: A survey, some research issues, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 602–622, 2016.

[71] Z. Zhu, P. Gupta, Q. Wang, S. Kalyanaraman, Y. Lin, H. Franke, and S. Sarangi, "Virtual base station pool: towards a wireless network cloud for radio access networks," in *Proceedings of the 8th ACM international conference on computing frontiers*. ACM, 2011, p. 34.

[72] G. Bhanage, I. Seskar, R. Mahindra, and D. Raychaudhuri, "Virtual basestation: Architecture for an open shared wimax framework," in *Proceedings of the second ACM SIGCOMM workshop on Virtualized infrastructure systems and architectures*. ACM, 2010, pp. 1–8.

[73] W. Kiess, P. Weitkemper, and A. Khan, "Base station virtualization for ofdm air interfaces with strict isolation," in *Communications Workshops (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 756–760.

[74] S. Costanzo, D. Xenakis, N. Passas, and L. Merakos, "Openb: a framework for virtualizing base stations in lte networks," in *Communications (ICC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3148–3153.

[75] K. Nakauchi, Y. Shoji, M. Ito, Z. Lei, Y. Kitatsuji, and H. Yokota, "Bring your own network?design and implementation of a virtualized wifi network," in *Consumer Communications and Networking Conference (CCNC), 2014 IEEE 11th*. IEEE, 2014, pp. 483–488.

[76] X. Xiaodong, Z. Huixin, D. Xun, H. Yanzhao, T. Xiaofeng, and Z. Ping, "Sdn based next generation mobile network with service slicing and trials," *China Communications*, vol. 11, no. 2, pp. 65–77, 2014.

[77] A. W. Dawson, M. K. Marina, and F. J. Garcia, "On the benefits of ran virtualisation in c-ran based mobile networks," in *Software Defined Networks (EWSDN), 2014 Third European Workshop on*. IEEE, 2014, pp. 103–108.

[78] X. Wang, S. Thota, M. Tornatore, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee, "Energy-efficient virtual base station formation in optical-access-enabled cloud-ran," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1130–1139, 2016.

[79] K. Amiri, M. Duarte, J. R. Cavallaro, C. Dick, R. Rao, and A. Sabharwal, "Fpga in wireless communications applications." Elsevier, Waltham, MA, 2012.

[80] J. Heinonen, P. Korja, T. Partti, H. Flinck, and P. Pöyhönen, "Mobility management enhancements for 5g low latency services," in *Communications Workshops (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 68–73.

[81] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network slicing based 5g and future mobile networks: mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[82] A. Ghosh, R. Ratasuk, I. Filipovich, J. Tan, and W. Xiao, "Random access design for umts air-interface evolution," in *2007 IEEE 65th Vehicular Technology Conference-VTC2007-Spring*. IEEE, 2007, pp. 1041–1045.

[83] S. Zarei, "Channel coding and link adaptation," *Ausgewhlte Kapitel der Nachrichtentechnik, WS*, vol. 2010, pp. 1–14, 2009.

[84] G.-Y. Lin, S.-R. Chang, and H.-Y. Wei, "Estimation and adaptation for bursty lte random access." *IEEE Trans. Vehicular Technology*, vol. 65, no. 4, pp. 2560–2577, 2016.

[85] A. Kukushkin, *Introduction to mobile network engineering*. Wiley Online Library, 2018.

[86] G. De la Roche, A. Alayón-Glazunov, and B. Allen, *LTE-advanced and next generation wireless networks: channel modelling and propagation*. John Wiley & Sons, 2012.

[87] M. Ulvan, R. Bestak, and A. Ulvan, "Ims signalling in ltebased femtocell network," in *UBICOMM 2010: The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2010.

[88] H. Fathi, S. S. Chakraborty, and R. Prasad, "On sip session setup delay for voip services over correlated fading channels," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 286–295, 2006.

[89] V. Bobrovs, S. Spolitis, and G. Ivanovs, "Latency causes and reduction in optical metro networks," in *SPIE OPTO*. International Society for Optics and Photonics, 2013, pp. 90 080C–90 080C.

[90] I. Al-Samman, A. Doufexi, and M. Beach, "A c-ran architecture for lte control signalling," in *Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd*. IEEE, 2016, pp. 1–5.

[91] E. Gelenbe, G. Pujolle, and J. Nelson, *Introduction to queueing networks*. Wiley New York, 1998.

[92] M. Panda, H. L. Vu, M. Mandjes, and S. R. Pokhrel, "Performance analysis of tcp newreno over a cellular last-mile: Buffer and channel losses," *IEEE Transactions on Mobile Computing*, vol. 14, no. 8, pp. 1629–1643, 2015.

[93] H.-L. To, S.-H. Lee, and W.-J. Hwang, "A burst loss probability model with impatient customer feature for optical burst switching networks," *International Journal of Communication Systems*, vol. 28, no. 11, pp. 1729–1740, 2015.

[94] B. Lampson, V. Srinivasan, and G. Varghese, "Ip lookups using multiway and multicolumn search," *IEEE/ACM Transactions on Networking (TON)*, vol. 7, no. 3, pp. 324–334, 1999.

[95] J. Hyun, J. Li, H. Kim, J.-H. Yoo, and J. W.-K. Hong, "Ipv4 and ipv6 performance comparison in ipv6 lte network," in *Network Operations and Management Symposium (APNOMS), 2015 17th Asia-Pacific*. IEEE, 2015, pp. 145–150.

[96] C. Mugga, D. Sun, and D. Ilie, “Performance comparison of ipv6 multihoming and mobility protocols,” in *Thirteenth International Conference on Networks (ICN).* IARIA XPS Press, 2014.

[97] I. Kim, Y. Jung, and Y.-T. Kim, “Low latency proactive handover scheme for proxy mipv6 with mih,” *Challenges for next generation network operations and service management*, pp. 344–353, 2008.

[98] S. Gundavelli, “K. leung, v. devarapalli, k. chowdhury, and b. patil,?” *Proxy Mobile IPv6*, 2008.

[99] R. Moskowitz, P. Nikander, and P. Jokela, “and t. henderson," host identity protocol,” RFC 5201, April, Tech. Rep., 2008.

[100] E. Nordmark and M. Bagnulo, “Shim6: Level 3 multihoming shim protocol for ipv6,” Tech. Rep., 2009.

[101] H. Jung and S. J. Koh, “Mobile-oriented future internet (mofi): Architecture and protocols,” *Release*, vol. 2, no. 2, 2010.

[102] D. Farinacci, D. Lewis, D. Meyer, and V. Fuller, “The locator/id separation protocol (lisp),” 2013.

[103] M. Gohar and S.-J. Koh, “A distributed mobility control scheme in lisp networks,” *Wirel. Netw.*, vol. 20, no. 2, pp. 245–259, Feb. 2014. [Online]. Available: http://dx.doi.org/10.1007/s11276-013-0605-x

[104] H. Luo, Y. Qin, and H. Zhang, “A dht-based identifier-to-locator mapping approach for a scalable internet,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 12, pp. 1790–1802, 2009.

[105] R. Wang, H. Hu, and X. Yang, “Potentials and challenges of c-ran supporting multi-rats toward 5g mobile networks,” *IEEE Access*, vol. 2, pp. 1187–1195, 2014.

[106] G. Pujolle, *Software Networks: Virtualization, SDN, 5G, Security.* John Wiley & Sons, 2015.

[107] S. Yan, X. Huang, M. Ma, P. Zhang, and Y. Ma, “A novel efficient address mutation scheme for ipv6 networks,” *IEEE Access*, vol. 5, pp. 7724–7736, 2017.

[108] S. Barré, J. Ronan, and O. Bonaventure, “Implementation and evaluation of the shim6 protocol in the linux kernel,” *Computer Communications*, vol. 34, no. 14, pp. 1685–1695, 2011.

[109] J.-I. Kim, H. Jung, and S.-J. Koh, “Mobile oriented future internet (mofi): Architectural design and implementations,” *ETRI Journal*, vol. 35, no. 4, pp. 666–676, 2013.

[110] M. Menth, D. Klein, and M. Hartmann, “Improvements to lisp mobile node,” in *Teletraffic Congress (ITC), 2010 22nd International.* IEEE, 2010, pp. 1–8.

[111] M. Gohar and S. J. Koh, “Network-based distributed mobility control in localized mobile lisp networks,” *IEEE Communications Letters*, vol. 16, no. 1, pp. 104–107, 2012.

[112] C. White, D. Lewis, D. Meyer, and D. Farinacci, “Lisp mobile node,” IETF Internet draft, Oct. 2011.[Online]. Available: http://tools. ietf. org/html/draft-meyer-lisp-mn-06, Tech. Rep., 2014.

[113] R. Escriva, B. Wong, and E. G. Sirer, “Hyperdex: A distributed, searchable key-value store,” in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication.* ACM, 2012, pp. 25–36.

[114] S. K. Afshar, N. Conley, K. Kiser, W. J. Leighton III, D. N. Lokhande, P. E. Mccrink, S. Neshatfar, B. J. Olszowy, R. Patel, S. Rajamannar *et al.*, “Paradigm in multimedia services creation methodology, and new service creation and service execution enviroments,” Nov. 22 2016, uS Patent 9,501,266.

[115] N. McKeown, “Software-defined networking,” *INFOCOM keynote talk*, vol. 17, no. 2, pp. 30–32, 2009.

[116] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “Openflow: enabling innovation in campus networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[117] M. Idri, “Mobility management based sdn-ipv6 routing header,” in *Software Defined Systems (SDS), 2017 Fourth International Conference on*. IEEE, 2017, pp. 150–155.

[118] J. Sen, “Mobility and handoff management in wireless networks,” *arXiv preprint arXiv:1011.1956*, 2010.

[119] R. Atkinson, S. Bhatti, and S. Hailes, “Evolving the internet architecture through naming,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 8, pp. 1319–1325, 2010.

[120] K. De Turck, E. De Cuypere, S. Wittevrongel, and D. Fiems, “Algorithmic approach to series expansions around transient markov chains with applications to paired queuing systems,” in *Performance Evaluation Methodologies and Tools (VALUETOOLS), 2012 6th International Conference on*. IEEE, 2012, pp. 38–44.

[121] M. Harchol-Balter, *Performance modeling and design of computer systems: queueing theory in action*. Cambridge University Press, 2013.

[122] D. Johnson, C. Perkins, and J. Arkko, “Mobility support in ipv6. ietf, request for comments: 3775, june 2004,” 2004.

[123] R. Koodli, “Fast handovers for mobile ipv6,” 2005.

[124] H. Soliman, C. Castelluccia, K. El Malki, and L. Bellier, “Hierarchical mobile ipv6 mobility management (hmipv6),” Tech. Rep., 2005.

[125] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, “Proxy mobile ipv6,” Tech. Rep., 2008.

[126] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia, “Fast handovers for proxy mobile ipv6,” Tech. Rep., 2010.

[127] J.-H. Lee, J.-M. Bonnin, I. You, and T.-M. Chung, “Comparative handover performance analysis of ipv6 mobility management protocols,” *IEEE Transactions on Industrial Electronics*, vol. 60, no. 3, pp. 1077–1088, 2013.

[128] H. Modares, A. Moravejosharieh, J. Lloret, and R. B. Salleh, “A survey on proxy mobile ipv6 handover,” *IEEE Systems Journal*, vol. 10, no. 1, pp. 208–217, 2016.

[129] W. Luo, X. Fang, M. Cheng, and Y. Zhao, “Efficient multiple-group multiple-antenna (mgma) scheme for high-speed railway viaducts,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 6, pp. 2558–2569, 2013.

[130] J. Kim, H.-S. Chung, I. G. Kim, H. Lee, and M. S. Lee, “A study on millimeter-wave beamforming for high-speed train communication,” in *Information and Communication Technology Convergence (ICTC), 2015 International Conference on*. IEEE, 2015, pp. 1190–1193.

[131] S. Banerjee, M. Hempel, and H. Sharif, "Decoupled u/c plane architecture for hetnets and high speed mobility: research directions & challenges," in *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2017, pp. 1–6.

[132] P. Dong, T. Zheng, S. Yu, H. Zhang, and X. Yan, "Enhancing vehicular communication using 5g-enabled smart collaborative networking," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 72–79, 2017.

[133] L. Lei, J. Lu, Y. Jiang, X. S. Shen, Y. Li, Z. Zhong, and C. Lin, "Stochastic delay analysis for train control services in next-generation high-speed railway communications system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 48–64, 2016.

[134] K. Banerjee, S. M. R. Islam, Z. I. Tahasin, and R. Uddin, "An efficient handover scheme for pmipv6 in ieee 802.16/wimax network," *International Journal of Electrical & Computer Sciences IJECS-IJENS*, vol. 11, no. 05, pp. 8–16, 2011.

[135] C. Makaya and S. Pierre, "An analytical framework for performance evaluation of ipv6-based mobility management protocols," *IEEE Transactions on wireless communications*, vol. 7, no. 3, pp. 972–983, 2008.

[136] M.-S. Chiang, C.-M. Huang, D.-T. Dao, and B.-C. Pham, "The backward fast media independent handover for proxy mobile ipv6 control scheme (bfmih-pmipv6) over heterogeneous wireless mobile networks." *J. Inf. Sci. Eng.*, vol. 34, no. 3, pp. 765–780, 2018.

[137] X. Li, R. Ferdous, C. F. Chiasserini, C. E. Casetti, F. Moscatelli, G. Landi, R. Casellas, K. Sakaguchi, S. B. Chundrigar, R. Vilalta *et al.*, "Novel resource and energy management for 5g integrated backhaul/fronthaul (5g-crosshaul)," in *Communications Workshops (ICC Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 778–784.

[138] X. Wang, G. Wang, R. Fan, and B. Ai, "Channel estimation with expectation maximization and historical information based basis expansion model for wireless communication systems on high speed railways," *IEEE Access*, vol. 6, pp. 72–80, 2018.

[139] F. Anwar, M. H. Masud, S. Bari, and S. A. Latif, "Enhanced handoff latency reduction mechanism in layer 2 and layer 3 of mobile ipv6 (mipv6) network," *Australian Journal of Basic and Applied Sciences*, vol. 7, no. 6, pp. 658–671, 2013.

[140] A. I. Sulyman, A. T. Nassar, M. K. Samimi, G. R. MacCartney, T. S. Rappaport, and A. Alsanie, "Radio propagation path loss models for 5g cellular networks in the 28 ghz and 38 ghz millimeter-wave bands," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 78–86, 2014.

[141] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-ghz integrated circuits and systems for wireless communications," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1390–1436, 2011.

[142] K. Sakaguchi, E. M. Mohamed, H. Kusano, M. Mizukami, S. Miyamoto, R. E. Rezagah, K. Takinami, K. Takahashi, N. Shirakata, H. Peng *et al.*, "Millimeter-wave wireless lan and its extension toward 5g heterogeneous networks," *IEICE Transactions on Communications*, vol. 98, no. 10, pp. 1932–1948, 2015.

[143] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, 2013.

[144] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[145] T.-T. Tran, Y. Shin, and O.-S. Shin, “Overview of enabling technologies for 3gpp lte-advanced,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 54, 2012.

[146] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, “Mmwave massive-mimo-based wireless backhaul for the 5g ultra-dense network,” *IEEE Wireless Communications*, vol. 22, no. 5, pp. 13–21, 2015.

[147] D. Wang and C. H. Chan, “Multiband antenna for wifi and wigig communications,” *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 309–312, 2016.

[148] S. Wyne, K. Haneda, S. Ranvier, F. Tufvesson, and A. F. Molisch, “Beamforming effects on measured mm-wave channel characteristics,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3553–3559, 2011.

[149] T. Reshmi and K. Murugan, “Light weight cryptographic address generation (lw-cga) using system state entropy gathering for ipv6 based manets,” *China Communications*, vol. 14, no. 9, pp. 114–126, 2017.

[150] S. Hagen, *IPv6 essentials*. " O’Reilly Media, Inc.", 2006.

[151] X. Yin and X. Cheng, *Propagation channel characterization, parameter estimation, and modeling for wireless communications*. John Wiley & Sons, 2016.

[152] K. Phemius and M. Bouet, “Monitoring latency with openflow,” in *Network and Service Management (CNSM), 2013 9th International Conference on*. IEEE, 2013, pp. 122–125.