On Designing Network Elements and Mechanisms of Frugal 5G Architecture for Rural Broadband Access

A thesis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

by

Meghna Khaturia
(Roll No. 14307R003)

Under the guidance of:
Prof. Abhay Karandikar

Electrical Engineering Department
Indian Institute of Technology Bombay
Powai, Mumbai 400076

Dedicated to my beloved parents, Mrs. Sadhana and Mr. Kishore Khaturia and my sisters, Neha and Preet.

Thesis Approval

The Thesis entitled

On Designing Network Elements and Mechanisms of Frugal 5G Architecture for Rural Broadband Access

by

Meghna Khaturia
(Roll no. 14307R003)

is approved for the degree of

Doctor of Philosophy

Prof. Debabrata Das (Examiner)	Prof. Saravanan Vijaykumaran (Examiner)
Prof. Abhay Karandikar (Advisor)	Prof. Sridhar Iyer (Chairperson)

Date: 12th July 2021

Place: IIT Bombay, Mumbai

Declaration

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Meghna Khaturia

Roll No. 14307R003

Date: 12^{th} July 2021

Abstract

Today, we live in a hyper-connected world that is redefining how people communicate, collaborate and congregate. However, we have a long way to go, with around half of the global population still unconnected. The rural areas of developing countries host a majority of the unconnected global population, suggesting a wide rural-urban digital divide. This thesis investigates the main challenges/requirements regarding connectivity in rural areas and how they can be effectively addressed.

The relevant research works and existing technologies follow a piecemeal approach and are ineffective in comprehensively addressing the rural broadband problem. Towards this, we follow a bottom-up approach, wherein we propose an abstract network architecture, named *Frugal 5G*, based on the rural connectivity requirements. We propose a heterogeneous access network along with a fog node at the edge to manage the network. The heterogeneous access network comprises macro base station to provide a large coverage area, Wireless Local Area Networks (WLANs) to provide high-speed connectivity in village clusters, and middle mile network to backhaul WLAN access points to the fiber point of presence. We employ software defined networking and network function virtualization paradigms to design the fog node. The Frugal 5G network architecture specifies the desired network characteristics of a rural broadband network, i.e., i) unified multi-Radio Access Technology (RAT) access network, ii) unified interworking function at the access network to interface with core/external networks, iii) multi-operator resource sharing, iv) localized communication, and v) flexible fog deployment.

To take the Frugal 5G network architecture towards adoption, it is vital to test the proposal's feasibility by realizing it using existing technology standards. Since Fifth Generation (5G) is the upcoming cellular technology, we realize the Frugal 5G network characteristics using 3GPP 5G network and refer to the realized network as *5G-Flow*. We complement the access network of 3GPP 5G with OpenFlow based controller and switches. The enhancements that we propose in the 5G access network are minimal and software-based. The 5G-Flow network addresses Frugal 5G network characteristics such as a unified multi-RAT access network along with a unified interworking function. 5G-Flow network also enables flexible fog deployment. Moreover, 5G-Flow network decouples an end-user's communication with RAN from its communication with the core network. This feature enables end-user to use any access technology to connect with any core network, e.g., use 5G New Radio based access to communicate with Fourth Generation (4G) core network. To examine the performance of the 5G-Flow network, specifically dataflow management of the multi-RAT access network, we have developed a 5G multi-RAT system-level simulator in MATLAB. The throughput and delay performance results demonstrate that 5G-Flow offers a significant improvement over the 3GPP 5G network.

In this thesis, we also address certain essential aspects of designing the middle mile network. Efficient planning of the middle mile network plays a crucial role in implementing a rural broadband network. We propose a simulated annealing based algorithm that aims to maximize system throughput along with minimizing the number of hops and middle mile nodes in the network. The performance results reveal that the proposed algorithm outperforms the greedy approach in critical parameters such as minimum user throughput and maximum hop count. The Frugal 5G network enables resource sharing among multiple operators. Advancing this concept, we formulate the problem of spectrum sharing among middle mile networks. We propose a graph-theoretic algorithm that allocates a combination of shared and dedicated channels to a base station, using carrier aggregation and listen before talk mechanisms. System-level simulations in ns-3 show that the proposed algorithm improves spectral efficiency as well as system fairness.

Standardization has played a pivotal role in taking any technology from conception to industrial adoption. Therefore, we have actively participated in standardization activities of P2061 standards development project, initiated by IEEE. P2061 aims to develop an "Architecture for Low Mobility Energy Efficient Network for Affordable Broadband Access". A major part of this thesis has been presented as contributions in P2061 meetings and may form part of the standard when it is finalized.

Contents

Acknowledgments

The journey of getting a doctoral degree has been one of the most rewarding times of my life, both professionally and personally. I have gathered some beautiful experiences and learnings for which I'll always be indebted to the people I met and interacted with in the past seven years. I will start by expressing my gratitude for my advisor Prof. Abhay Karandikar. His dedication, energetic persona, and research ethics have been genuinely inspiring. His expertise has been invaluable in formulating the problem and developing a practical solution for it. Apart from being a very supportive advisor, he has been a very kind and caring mentor too.

I sincerely appreciate Prof. Kumar Appaiah for always being there whenever I needed help. I am grateful to him for sharing his wisdom and experiences and for being very encouraging.

I would like to extend my heartfelt thanks to Mr. Pranav Jha. He has been extremely patient in making me understand the subject and has tirelessly answered every question of mine. I have learned so much from the numerous discussions with him on topics related to research as well as on matters beyond research. He has been a very kind and considerate mentor and has inspired me to be kind as well.

Next, I want to thank my Research Progress Committee (RPC) members, Prof. Saravana Vijaykumaran and Prof. Prasanna Chaporkar, for their valuable suggestions. The feedback provided by them has been really helpful in improving the quality of the thesis. I also want to thank Prof. Chaporkar for the engaging discussions with all the Ph.D. students in the lab during presentations. I would also like to thank the IIT professors who provide such a warm and compassionate environment for learning. Through the courses I took in IIT, I genuinely started to enjoy the process of learning.

My special thanks to Sonal Ma'am, Sangeeta Ma'am, Beena, Margaret, Aditya ji, and Rajesh ji for their tremendous support with all the paperwork and logistics. Without their support, the journey wouldn't have been this smooth. I really appreciate the assistance provided by EE office staff, Mr. Santosh, and Mrs. Madhu.

I have made some very special friends on this journey, and I will forever be indebted to them for making IIT feel like home. Sadaf, Shuvolina, Sweety, and Rashmi ma'am have been by my side throughout this journey. Without their kindness and encouragement, I wouldn't have been able to get through this journey. I would also like to thank my dear friends and lab mates Sarbani Ma'am, Akshatha, Indu, Shashi, Sagar, Shoumik, Gaurang, Shubham, Arghyadip, Pradnya, Annapurna, and Tejashri. Personally and professionally, I have learned and grown so much because of them. I would also like to thank Annu, Utkarsh, Megha, Nandan, Uma, Prakash, and Richa for their unconditional love and support.

I am eternally grateful to my loving and caring family. My parents have always supported me with my decisions and stood by me through thick and thin. Their unwavering support has given me the courage to leave my job and pursue a master's and then a doctorate degree. I am thankful to my sisters (Neha and Preet), my brother-in-law (Kamal), and my nieces (Saanvi and Divisha) for their unconditional love and support, which has been an incredible source of strength for me.

Meghna Khaturia
July 2021

List of Tables

List of Figures

List of Abbreviations

3G	Third Generation
3GPP	Third Generation Partnership Protocol
4G	Fourth Generation
5G	Fifth Generation
5GC	5G Core Network
AMF	Access and Mobility Management Function
AP	Access Point
API	Application Programming Interface
ATSSS	Access Traffic Steering, Switching & Splitting
ARPU	Average Revenue Per User
BS	Base Station
BTS	Base Transceiver System
CA	Carrier Aggregation
CAPEX	Capital Expenditure
CAPWAP	Control And Provisioning of Wireless Access Points
CPE	Customer Premises Equipment
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
DHCP	Dynamic Host Configuration Protocol
DSL	Digital Subscriber Line
DTMC	Discrete Time Markov Chain

DRB	Data Radio Bearer
eMBB	enhanced Mobile Broadband
eNB	enhanced NodeB
EIRP	Effective Isotropically Radiated Power
ETSI	European Telecommunications Standards Institute
F1-AP	F1-Application Protocol
FCCA	Fairness Constrained Channel Allocation
gNB	Next-generation NodeB
GP	Gram Panchayat
GRE	Generic Routing Encapsulation
GTP	GPRS Tunneling Protocol
IFS	Inter Frame Space
IMT-2020	International Mobile Telecommunications-2020
IN	Intermediate Node
IPsec	IP Security
ISM	Industrial, Scientific and Medical
ISD	Inter-Site Distance
ITU	International Telecommunication Union
LBT	Listen Before Talk
LDCs	Least Developed Countries
LMLC	Low Mobility Large Cell
LoS	Line of Sight
LTE	Long Term Evolution
LWA	LTE-Wireless Local Area Network Aggregation
MAC	Media Access Control
MANO	Management and Orchestration

MATLAB	Matrix Laboratory
MLMF	Mobility and Load Management Function
MME	Mobility Management Entity
MRN	Multi-RAT Network
N3IWF	Non-3GPP Interworking Function
NAS	Non-Access Stratum
NETCONF	Network Configuration Protocol
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure
NGAP	Next-Generation Application Protocol
NLoS	Non-Line of Sight
NR	New Radio
ns-3	network simulator-3
OF-Config	OpenFlow-Configuration
OPEX	Operational Expenditure
PDU	Protocol Data Unit
PE	Provider Edge
PCRF	Policy and Charging Rules Function
PoP	Point of Presence
QoE	Quality of Experience
QoS	Quality of Service
RACF	RAT Agnostic Control Function
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RF	Radio Frequency

RMa	Rural Macro
RoFSO	Radio on Free Space Optics
RoW	Right of Way
RRC	Radio Resource Control
RSCF	RAT Specific Control Function
SA	Simulated Annealing
SCTP	Stream Control Transmission Protocol
SDN	Software Defined Networking
SDAP	Service Data Adaptation Protocol
SM	Spectrum Manager
SMF	Session Management Function
SNR	Signaling to Noise Ratio
SRB	Signaling Radio Bearer
SSID	Service Set Identifier
TCP/IP	Transport Control Protocol/Internet Protocol
TNGF	Trusted Non-3GPP Gateway Function
TPM	Transition Probability Matrix
TPR	Technical Performance Requirements
TSDSI	Telecommunications Standards Development Society, India
UE	User Equipment
UDF	User Database Function
UDP	User Datagram Protocol
UHF	Ultra High Frequency
UMi	Urban Micro
UPF	User Plane Function
VAP	Virtual Access Point

VNF Virtual Network Function

VoIP Voice over IP

W-AGF Wireline-Access Gateway Function

WiMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

Chapter 1

Introduction

"We shape our tools and, thereafter, our tools shape us." — John Culkin [1].

The Internet, as we know it today, was invented with the broader adoption of Transport Control Protocol/Internet Protocol (TCP/IP); subsequently the Internet has changed and shaped our world in the most unimaginable ways possible. The Internet has democratized time, space, resources, knowledge, and ideas; it has literally broken the narrow domestic walls. With the advancement in broadband technology and its increased availability, the world has transformed into a "global village" [2], it has become "flat" [3]. The communities across the globe are communicating, collaborating, co-creating, and commuting together for a sustainable world, and the Internet connectivity provides *highways* for this journey. The COVID-19 pandemic, one of the largest public health crisis since the Spanish Flu (1918), would have disrupted the entire world order had it not been for the Internet. From COVID-19 contact tracing to remote working to contactless payment has been possible only due to high-speed Internet availability. Access to the Internet is a key enabler for democracy, without which values such as justice, equity, and access are difficult to imagine in modern times. Therefore, the United Nations has recommended "Right to Internet Access" to be a fundamental right of citizens globally [4].

1.1 The State of Broadband Connectivity

According to a report by the International Telecommunication Union (ITU), around 49% of the global population is unconnected [5]. These statistics suggest that a significant part

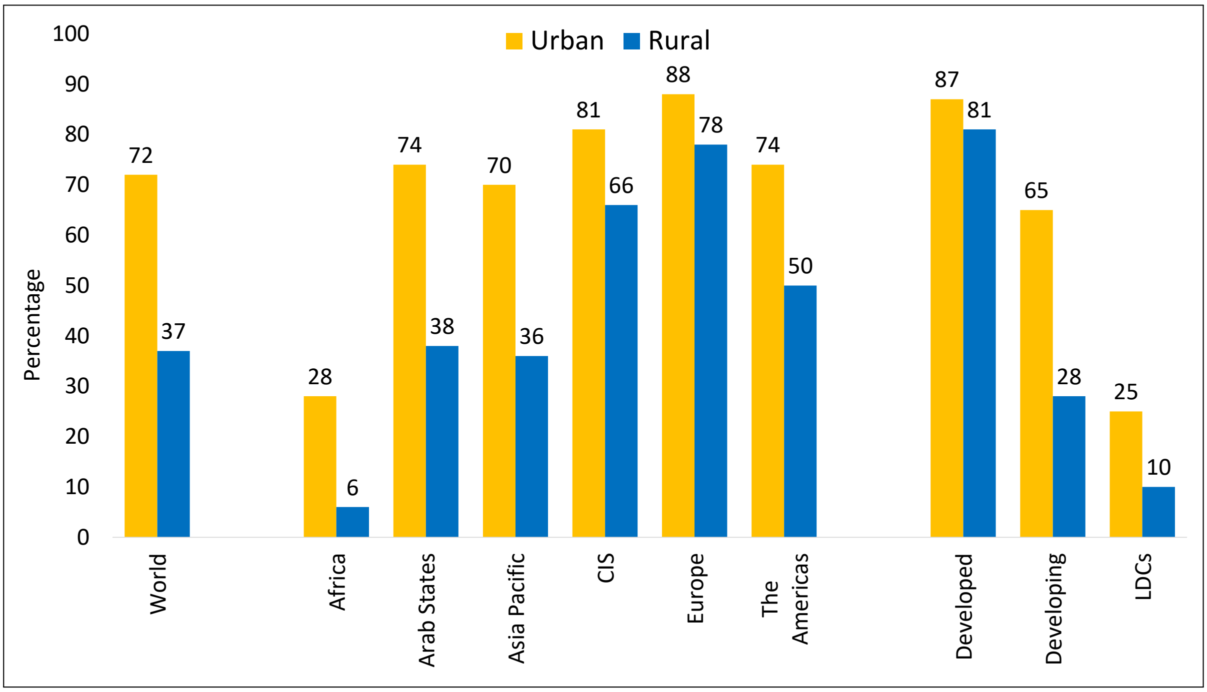


Figure 1.1: Percentage of households with Internet access, 2019 (ITU) [5]

of the world population is unconnected and cannot experience the benefits that broadband Internet can provide. While the majority has access to the Internet in developed countries, only 44% of individuals in developing countries and 19% of individuals in the Least Developed Countries (LDCs) are using the Internet. Moreover, there are regional differences in datarate usage per Internet user. While the average datarate usage for individuals in developed countries is 189 kbit/s per Internet user, the same is 91 kbit/s and 21 kbit/s per Internet user in developing countries and LDCs, respectively [5]. On further analysis, we observe that the adoption of the Internet is unevenly distributed across urban and rural areas in developing countries and LDCs, giving rise to a significant rural-urban digital divide. On a global scale, the percentage of households with Internet access at home in urban areas is approximately double that of rural areas, as shown in Figure 1.1. In a developing country like India, there are 65 broadband subscriptions per 100 individuals in urban areas while the same is 29 in rural areas, as indicated in Figure 1.2 [6]. A similar rural-urban digital divide can be observed in developing countries such as China, Nigeria, Ethiopia, Bangladesh, and Tanzania [7].

Having established the inadequate availability of broadband connectivity in rural areas (of developing countries and LDCs), it would be an understatement to say that the

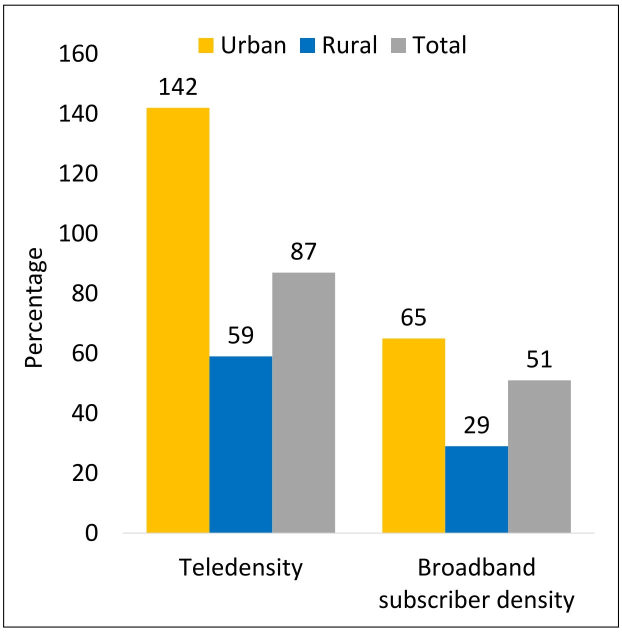


Figure 1.2: Teledensity and broadband subscriber density of rural and urban areas in India (as on 31st March 2020) [6]

rural areas are at a disadvantage. Absence of Internet has profound implications on the lives of the unconnected population. Besides lacking access to information, social interaction and better trade opportunities, they can't share their voices, ideas and contributions with the global community. With minimal facilities in most hospitals, schools, and offices in rural areas, the availability of rural broadband can make a real difference. Advancements in the field of virtual reality has now made it possible to simulate real workspaces for the purposes of distance education, skill development and telemedicine. Moreover, digitization of rural areas can help in effectively delivering farmer extension services [8] which will enhance their knowledge and improve their technical skills. Farmers could be exposed to new scientific research, farming methods and information regarding new plant and animal diseases. This will not only increase the farmers' income but also increase the agriculture efficiency. Digitization is also a driver for increased employment opportunities, leading to improved living standards in rural areas.

The impediments to enabling rural connectivity are quite complex. An initial investigation into the underlying reasons for poor availability of broadband in rural areas suggests challenges such as sparse population density, intermittent supply of electricity, low-income resident population, and difficult terrain. Although there are various existing technologies and research proposals that aim to overcome these challenges, the rural connectivity problem is far from solved (as suggested by the recent broadband statistics discussed above). Therefore, it is imperative to analyze these existing solutions to understand their limitations and pave the road ahead for enabling rural connectivity.

1.2 Existing Technologies for a Rural Broadband Network: Strengths and Limitations

Conventionally, a communication network comprises the core, the backhaul, and the access network, as shown in Figure 1.3. The *core network* acts as a gateway to connect with other networks. It consists of servers, routers and switches that are capable of handling large amount of traffic from plurality of users. It also handles authentication, aggregation, switching, facilitation of network services, and charging. The *backhaul network* extends the core network connectivity to a designated point referred to as Point of Presence (PoP) within a neighborhood. A high-speed backhaul network with sufficient redundancy is essential for providing reliable connectivity. The *access network* is the portion of the network that extends the backhaul connectivity to the end-user through various technologies. On closer observation, we see that the access network consists of two components a) a distribution network that extends connectivity from PoP to end-user premises (such as an office building or a residential area) and b) a network that provides last mile connectivity to the end-user. For the purpose of this thesis, we refer to the former component as *middle mile network* and the latter as *last mile network*, as depicted in Figure 1.3.

Now, we try to understand which components of a communication network are bottlenecks for enabling connectivity in rural areas. The importance of fiber infrastructure in the dissemination of backhaul network can not be understated. In developed countries, fiber infrastructure is widely deployed across all locations and hence high-speed broadband network distribution is viable via fiber. In contrast, the developing countries lack adequate fiber infrastructure, which is further underscored by the fiber deployment-to-population ratio (in km). While this ratio is 0.1 in India, it is 1.1 and 1.7 in China and US, respectively [9]. Therefore, it is essential to improve fiber infrastructure in developing countries. Towards this, the Government of India, under its Digital India program, has

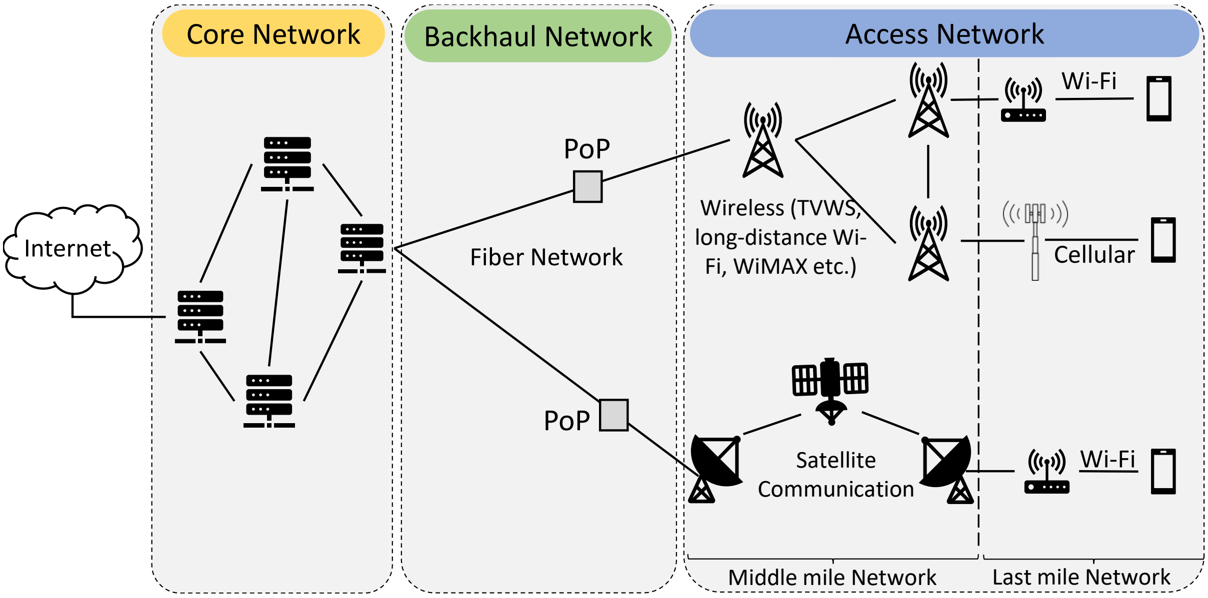


Figure 1.3: A communication network and its components

initiated a project called “BharatNet” [10]. This project aims at providing a PoP with optical connectivity to local self-government office at the village level called Gram Panchayat (GP) in India. The villages are generally located within a few kilometers of the GP office and remain unconnected if fiber connectivity reaches only GP office. A similar pattern is observed in most developing countries, wherein the fiber connectivity reaches only at designated points in the vicinity of the rural areas (PoPs), leaving a large part unserved.

Extending connectivity from PoPs to end-users in an efficient manner is a critical challenge that impedes rural connectivity. Therefore, in this thesis, we mainly focus on access network technologies suitable for the rural areas. We analyze the strengths and limitations of existing middle mile as well as last mile technologies (each being a component of the access network).

1.2.1 Technologies for Middle Mile Network

Multiple technologies can be used in the middle mile network as shown in Figure 1.3. Deployment of fixed broadband infrastructure such as optical fiber or Digital Subscriber Line (DSL) for the middle mile network can be very time consuming and expensive due to Right of Way (RoW) challenges and the cost of civil works. Even the management of wired networks post deployment in a rural/remote location is very challenging. Therefore, we mainly focus on wireless technologies which cost-effective, easy to deploy and easy to maintain and these are discussed below.

Long Distance Wi-Fi Networks

In the last decade, there have been several initiatives around the world to provide broadband connectivity to rural communities via long-distance IEEE 802.11 (Wi-Fi) based networks [11]. In [12], the authors have deployed two rural testbeds based on long-distance Wi-Fi networks viz. digital Gangetic Plains and Ashwini. In [13], the authors have proposed enhancements for Wi-Fi-based long distance networks to improve the shortcomings of these networks, such as packet losses and low link utilization. Hop-Scotch is a long-distance Wi-Fi network based rural testbed in UK [14]. Furthermore, there has been substantial research interest in designing the long-distance Wi-Fi based mesh networks for rural areas [15–17]. In [18], the authors address an essential issue of planning the long-distance Wi-Fi mesh network in rural areas to minimize the cost of network deployment. The authors in [19] have proposed a novel channel allocation algorithm for a rural wireless mesh network that allows the links to operate in full-duplex mode. Hence, significant throughput gains are observed. In addition to planning the long-distance Wi-Fi networks, reuse of the available infrastructure is also crucial to realize an effective solution.

Despite many efforts listed above, certain gaps in these approaches prevent them from providing a reliable solution to the problem. Wi-Fi operates in Industrial, Scientific and Medical (ISM) bands, i.e., 2.4 GHz or 5.8 GHz frequency bands. When working with these frequency bands over a long distance point to point link, a clear Fresnel zone and Line of Sight (LoS) are required between transmitter and receiver. Therefore, towers of sufficient height and high gain directional antennas are necessary infrastructure requirements for establishing the link. Moreover, the maintenance of such towers is challenging in remote areas. Hence, the overall cost of the network increases significantly, despite Wi-Fi devices being highly commoditized. The unlicensed band has other problems such as Effective Isotropically Radiated Power (EIRP) limit (e.g., 4 W in India for 2.4 GHz band [20]). Covering longer distances along with support for high throughput may become a difficult task with such EIRP limits.

TV White Spaces

TV White Spaces are the unutilized frequency bands allocated for broadcasting services. These white spaces exist primarily due to the switchover from analog to digital television, leaving large chunks of spectrum unused. In India, the TV UHF band (470-585 MHz) is highly underutilized [21]. Around 12 out of 15, channels in the TV UHF band (470-585 MHz) are available at any given location in India. Since the TV UHF band has good propagation characteristics, it has a significant potential to provide broadband services in rural areas. Several TV White Space standards have been developed in the past, such as IEEE 802.22, IEEE 802.11af, IEEE 802.19.1, IEEE 802.15.4m, IEEE 1900.7, and ECMA 392 [22]. The two most relevant standards providing rural connectivity are IEEE 802.22 and IEEE 802.11af, and are discussed next.

IEEE 802.22 has been developed to access the TV White Spaces using cognitive techniques [23]. The most important application of this standard is to provide wireless broadband access in rural and remote areas. The standard enables a large range of 33 km with only 4 W of EIRP. The standard uses both geolocation database and sensing-based techniques to access the TV White Spaces. Although this technology presents a promising solution, IEEE 802.22 is commercially available from a few vendors only, making IEEE 802.22 technology unaffordable. Also, considering the TV White Space scenario in India or any developing country where large parts of the band are unutilized, cognitive capabilities are not essential.

IEEE 802.11af standard (White-Fi) has been formulated to adapt the existing IEEE 802.11 for TV band operation [24]. IEEE 802.11af systems operate on frequencies below 1 GHz and use geolocation database to access the TV band. This standard has been designed to address congestion in the unlicensed bands, i.e., 2.4 GHz and 5 GHz spectrum. There are two operating scenarios of IEEE 802.11af viz. indoor and outdoor. The indoor scenario has a range of up to 100 m (similar to Wi-Fi), whereas the outdoor scenario has a range of about a few kilometers and is more suited for the rural setting. Although this technology can be made cost-effective (due to highly commoditized IEEE 802.11 devices), its spectral efficiency is low. Therefore, high data rates cannot be supported using these devices.

Free Space Optics

Free space optical networks use free space to transmit and receive data via an optical link. As this technology employs visible light for communication, there is a significant advantage of zero license fees. Free space optical links can support Gbps data rates. Moreover, optical components are generally less expensive and power efficient. However, there are certain challenges facing the optical network deployment. The optical networks are susceptible to atmospheric turbulence, which hampers the performance of the optical link. Besides, the optical network uses a narrow beam, which requires precise alignment to establish a successful link. In [25], the authors propose a network design concept based on Radio on Free Space Optics (RoFSO). The authors also provide experimental results obtained from the performance evaluation of RoFSO. The RoFSO system performed well in suitable weather conditions but is susceptible to rainfall.

Satellite communication

Satellite communication is suitable for regions that are quite remote, wherein laying optical fiber is not a suitable option. However, satellite connectivity is expensive and is not the best option for low-income areas. Moreover, the signal experiences significant delays due to considerable distance between transmitter and receiver. In [26], the authors enable Internet connectivity in the rural areas of Zambia via VSAT satellite connectivity and local Wireless Local Area Network (WLAN) network. They provide useful insights regarding the problems encountered when enabling rural connectivity, such as power instability and its impact on electrical equipment, high satellite connectivity costs, absence of local hardware suppliers, and sensitivity to weather.

Other Potential Technologies

Similar to long-distance Wi-Fi networks, *Microwave* based middle mile network can be employed to extend connectivity to rural areas via a radio link. Microwave network uses licensed frequency bands; therefore, interference is not a critical challenge. However, the key limitations are the cost of erecting and maintaining tall towers and the spectrum licensing fee.

Based on 60 GHz *mmWave technology*, Terragraph, a multi-hop wireless network, has been proposed by Facebook Connectivity Lab [27]. Terragraph uses lamp posts and rooftops of buildings to form a multi-hop network connecting the PoP to the access points. Therefore, it is suitable only in rural areas with a large population and the availability of required infrastructure such as lamp posts and buildings.

Worldwide Interoperability for Microwave Access (WiMAX) is a wireless communication standard that has been designed to work in metropolitan areas [28]. It provides coverage of up to 48 km for fixed stations with a data rate of 70 Mbps and coverage of $5 - 15$ km for mobile stations with a data rate of up to 15 Mbps. It supports multiple topologies: fixed point to point, fixed point to multipoint, and mobile. The fixed WiMAX topology can be effectively used to solve the middle-mile problem in rural areas. However, the deployment cost is the main limiting factor for using WiMAX in these areas.

1.2.2 Technologies for Last Mile Network

Before discussing the last mile technologies, it is essential to identify the most commonly used last mile/hop technologies. We observe that around 53% of total Internet users access the Internet via their smartphones [29]. This percentage is even more significant in Africa and Asia, with 70% and 62% of users accessing the Internet via smartphones, respectively. These statistics indicate that users increasingly prefer smartphones over desktop/laptop to access the Internet. Moreover, in rural areas of Asia/Africa, the likelihood of users owning a smartphone to access the Internet is greater than owning a desktop/laptop, as smartphones are relatively less expensive. Therefore, we consider the most common access technologies available on smartphones (also available on desktops/laptops), viz., cellular technology, and Wi-Fi.

Cellular Technology

Since the fiber infrastructure is largely unavailable in developing countries, cellular connectivity has become a standard norm for providing broadband access. However, the Third Generation (3G) or Fourth Generation (4G) cellular connectivity, capable of providing broadband access, have a relatively light footprint in rural areas when compared to urban/sub-urban areas. This fact is further highlighted if we observe the radio coverage maps of the 4G cellular network owned by Jio Infocomm Limited in India [30]. The radio coverage is limited to the cities and the highways connecting them. Large parts of the country, particularly rural areas, do not have access to 4G connectivity. The underlying reason that the cellular operators cannot extend their networks in rural areas is that it is economically unviable for them. To deploy a cellular network, Capital Expenditure (CAPEX) is significant, comprising the cost of the Base Transceiver System (BTS), fiber deployment, and passive infrastructure. A considerable part of the Operational Expenditure (OPEX) includes fuel and energy costs. However, the subscriber base is low in rural areas due to the sparse population density. Therefore, the Average Revenue Per User (ARPU) is low, discouraging operators from investing in these areas.

The vision for Fifth Generation (5G) standards as specified by International Mobile Telecommunications-2020 (IMT-2020) primarily focuses on requirements such as peak download data rate of 20 Gbps, 1 ms latency, and support for user speed up to 500 km/h [31]. These requirements are more suited towards urban use-cases. The connectivity requirements of rural areas are quite different (as discussed in Section 1.4), and the existing 5G standards are unlikely to meet those requirements [32]. Despite all the limitations mentioned above, 5G networks have tremendous potential to enable broadband connectivity in rural areas due to its rapid expansion globally. Several studies in literature [33–35] propose modifications to the 5G network to meet the objectives of rural networks. In [34], the authors have examined the possibility of large cells and mounting Remote Radio Heads on the unmanned aerial vehicles. Economic analysis of these techniques show that the monthly subscription fee can be kept sufficiently low. In [33], the authors have analyzed the impact of beamforming and cell discontinuous transmission on energy consumption of the network. The analysis suggests that beamforming can be a key enabler to significantly reduce power consumption, thereby decreasing the operational cost. The work presented in [35] has proposed the usage of TV White Space spectrum with the 5G infrastructure for providing coverage in rural areas. An optimization problem has been formulated to minimize the capital and operational expenses of the proposed system. 5G and beyond mobile networks, if enhanced suitably for the rural areas, can prove to be a very promising solution.

Wi-Fi

The potential of fixed wireless technologies such as Wi-Fi (IEEE 802.11) for last mile network has been discussed thoroughly in literature as well as realized in urban/suburban areas. Wi-Fi is widely used for enterprise/campus networks, public hotspots, and home networks. The Wi-Fi devices have been highly commoditized due to its wide-spread use, resulting in substantial cost-reduction. In rural areas, characterized by a sparse population, Wi-Fi based last mile network can be a potential solution. However, the major challenge lies in efficiently backhauling Wi-Fi access points. Wi-Fi, along with a wireless middle mile network designed for rural areas, can enable rural connectivity in an affordable manner.

1.2.3 Open Challenges

Following our analysis of various technologies discussed above, we understand that no single technology overcomes all the challenges concerning rural broadband connectivity. The cost of deployment and inability to effectively reach remote areas renders these technologies infeasible for rural areas. Moreover, these technologies have been primarily designed to serve urban use-cases and provide a piecemeal solution to address the rural broadband challenges.

One approach to address the rural connectivity problem can be designing and developing an entirely new technology specifically for rural areas. However, it is impractical to design a single technology applicable to rural areas with varied demographics, terrain, and infrastructure availability. Another approach is to identify and exploit the opportunities and trade-offs offered by rural areas and propose a solution that uses existing technologies. We adopt the latter approach, wherein we employ the existing middle mile and last mile technologies under a novel network architecture. By utilizing state-of-the-art network paradigms such as software-defined network and fog computing, the network architecture can be designed to suit rural use-cases. To enable industrial adoption of any network architecture, standardization plays a key role and must be considered.

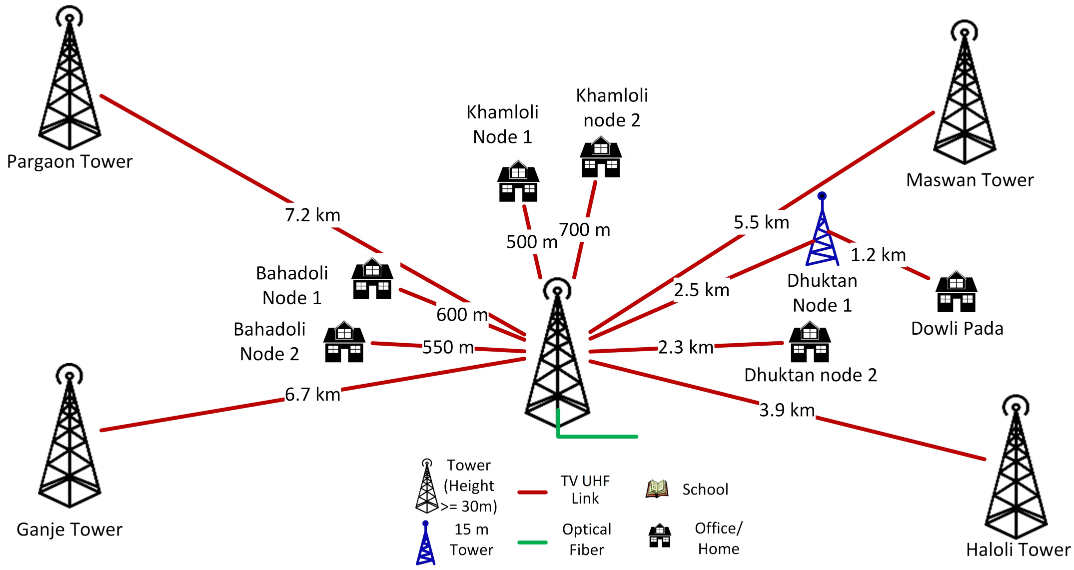
To understand the trade-offs and opportunities offered by rural areas, we briefly discuss the two large-scale rural broadband testbeds deployed as part of the rural broadband project at the Indian Institute of Technology (IIT) Bombay (of which the author was part of the team). We describe the objective, implementation, and results obtained from these testbeds. Further, we explain the rural connectivity challenges/characteristics through the insights that we gained from implementing the two rural testbeds.

1.3 Revisiting Palghar Testbeds

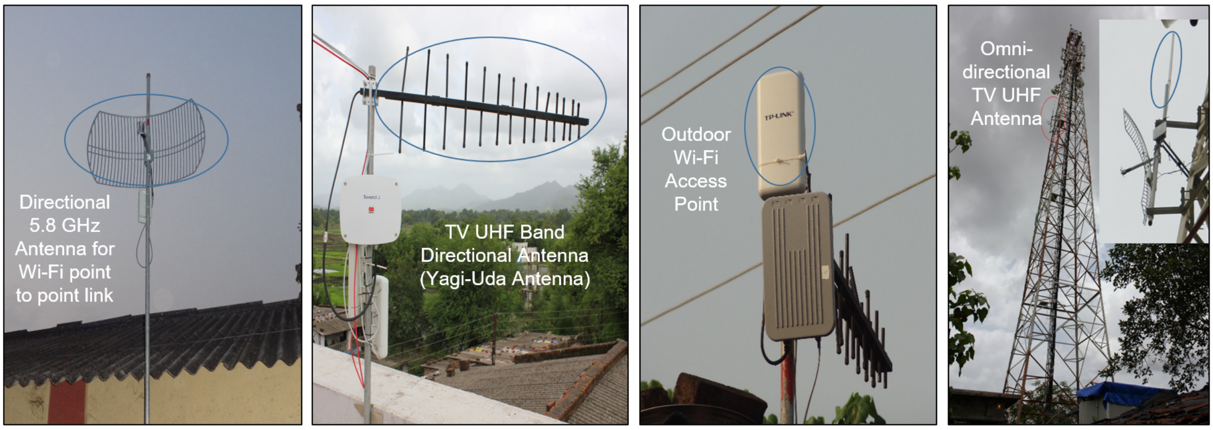
As part of the Rural Broadband Project at IIT Bombay, two large scale testbeds were set up in the rural areas of Palghar, Maharashtra, India [36–38]. The vision of these testbeds was to extend the broadband connectivity from the PoP (located several km away) to the villages. The last mile access to the villagers was provided by installing Wi-Fi Access Points (APs). To backhaul Wi-Fi APs to the PoP, a wireless multi-hop middle mile network was deployed. The middle mile network comprises middle mile AP located near the PoP and middle mile clients located in the villages.

As shown in Figure 1.4, the first testbed was deployed to analyze the feasibility of using the TV Ultra High Frequency (UHF) band to efficiently implement a middle mile network, providing high-speed broadband access in the villages. As discussed before, the TV UHF band has good propagation characteristics and is highly underutilized in India [21]. To implement a TV UHF based middle mile network in the testbed, we developed a low-cost TV UHF band radio using off the shelf IEEE 802.11g based devices and upconverting the Radio Frequency (RF) from 2.4 GHz to 500 MHz. This prototype enabled significant cost reductions in the testbed. Using this prototype, we deployed a point to multi-point middle mile network, as shown in Figure 1.4a. The TV UHF band testbed spanned an area of about 25 sq. km and covered seven villages in Palghar. Several LoS and Non-Line of Sight (NLoS) links were successfully established and analyzed in this testbed. Some results and testbed images are depicted in Figure 1.4. Further details concerning the testbed results can be found in [36].

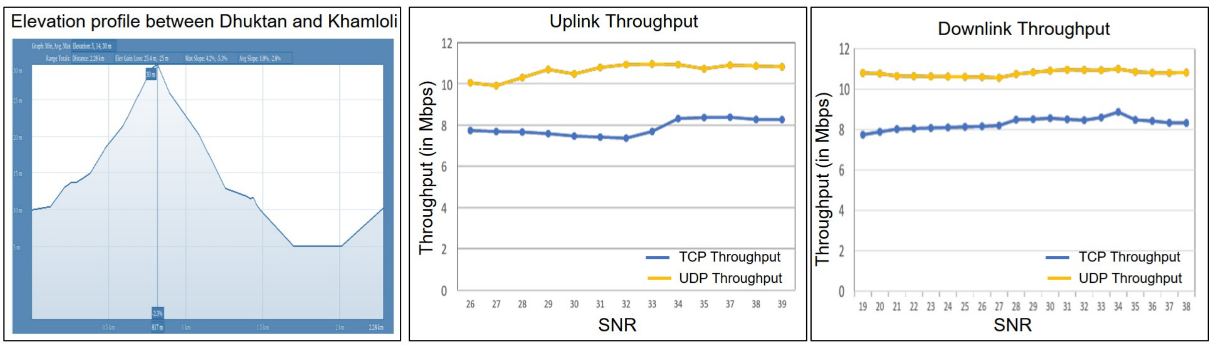
The second testbed was based on IEEE 802.11 (5.8 GHz). To get a detailed insight into developing a large scale solution for rural broadband, we envisioned an even larger scale testbed compared to the TV UHF band testbed deployed in Palghar. This necessitated the development of a software-based planning tool to design the network efficiently [10]. The planning tool considers bandwidth availability, tower availability, and throughput requirement based on population size. Based on these parameters, the tool determines wireless link feasibility. The testbed spanned 25 villages over an area of about 350 sq. km. There were 6 clusters across 25 villages where each cluster comprised 4-5 villages. The results of the testbed have been presented in [38].



(a) Network topology of Palghar testbed



(b) Images of TV UHF band and Wi-Fi devices from the TV UHF Band testbed in Palghar, Maharashtra India



(c) Elevation profile of Dhuktan-Khamloli link along with obtained TCP/User Datagram Protocol (UDP) throughput results are shown. The channel bandwidth was 10 MHz and transmit power was 27 dBm.

Figure 1.4: TV UHF band testbed

The planning, deployment, and experimentation of two large-scale testbeds in rural areas have provided us with insights concerning rural connectivity requirements. These insights are discussed next.

1.4 Rural Connectivity Requirements: Learnings From the Palghar Testbeds

This section presents the unique traits that characterize the rural/remote regions and the key requirements/challenges that impede the deployment of a broadband network there.

- *Affordability:* It has been observed that the average income of people living in rural areas is typically much less than that in urban areas. This income disparity is worse in developing countries. For example, the average monthly income of a rural household in India is nearly \$122 (₹8931) [39], less than 50% of an average urban household in the country. Therefore, affordability is a primary concern while developing a network framework.

- *Support for Clustered Service Provision:* Rural areas are populated in clusters that are generally far apart from each other with vast open areas in between, as shown in Figure 1.5. The size of clusters shown in the figure can vary from 0.01 sq. km to 0.5 sq. km. Based on the rural land area [40] and total number of villages in the India (650,000 [41]), there is approximately one village cluster every 3-4 sq. km. This implies that providing a similar type of services over a large geographical area may be economically inefficient. The network must be efficiently designed based on rural demographics. Heterogeneous access technologies can be used to provide different levels of services in a given area.

- *Support for Regional Content Generation and Storage:* Globally, 80% of online content is available in 1 of 10 languages which only 3 billion people speak [42]. In majority of rural areas, this 80% content has little or no relevance. Additionally, there exists a gap to facilitate content generation in regional languages for consumption.

- *Low Mobility:* People in rural areas, especially in developing countries such as India, are either pedestrian or move at low speeds, typically lower than 50 km/h. In

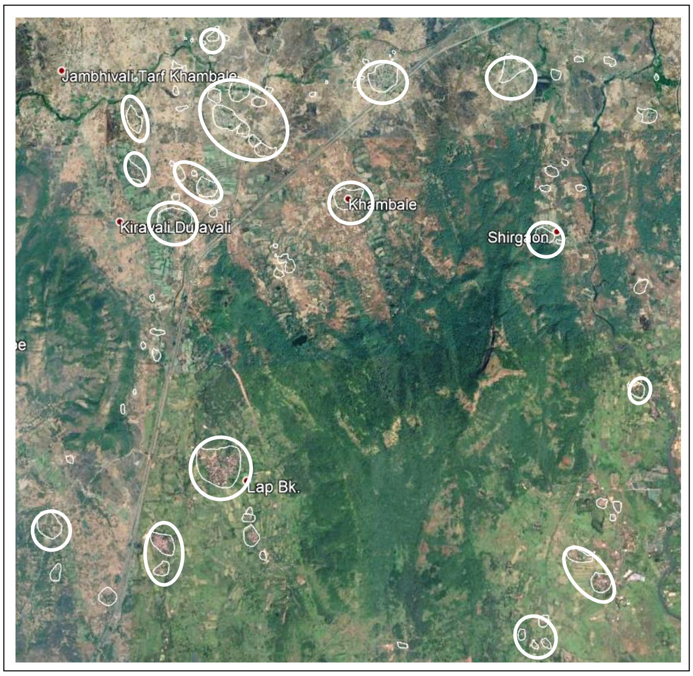


Figure 1.5: An example of rural area (approx. 50 sq. km) depicting the clustered settlement, wherein an ellipse denotes a habitat

such a scenario, supporting high-speed mobility (up to 500 km/h) may not have much relevance. Support for high mobility poses various challenges on the design of a cellular system, such as developing highly complex channel estimation and synchronization techniques [43].

- *Remote/Inaccessible Areas:* Many parts of the rural areas may be remote and not easily accessible. The maintainability and availability of the network equipment deployed in such remote regions may be of significant concern. Efficiently planned wireless middle mile network that extends in these remote regions can address the challenges of accessibility and maintenance.

- *Support for Local Communication*: One of the crucial observations from our Palghar testbed is that the people in rural areas have a strong community bond. This observation suggests that their communication pattern has a distinct localized nature, and a significant fraction of the communication, especially peer-to-peer communication, is between people living in close vicinity of each other. Localized support in the network can be extremely valuable in rural areas when the backhaul connectivity is disrupted.

- *Support for Renewable Energy Sources*: In rural areas of developing countries such as India, the availability of electricity from the grid is intermittent in nature [44]. Therefore, it is impossible to guarantee a continuous Internet service if the network equipment is solely dependent on the electric grid. Therefore, the usage of renewable energy sources such as solar power is essential.

1.5 Contributions and Roadmap

The aim of this thesis is to propose solutions that address the rural connectivity requirements stated above. The thesis is organized into seven chapters. Chapter 2-6 present the contributions of thesis. We now outline the chapter wise contributions below.

- In Chapter 2, we develop an abstract network architecture following the rural connectivity challenges highlighted in Section 1.4. We refer to the proposed network architecture as *Frugal 5G*. We propose a heterogeneous access network comprising macro Base Station (BS) (for providing basic connectivity) and WLANs (for providing high-speed connectivity in clusters). The WLANs are backhauled via wireless multi-hop middle mile network. To manage the proposed heterogeneous access network, we proposed a fog node at the edge and employ state-of-the-art network paradigms such as software-defined networking and network function virtualization to design it. The fog node presents a unified multiple access framework that integrates multiple last mile and middle mile technologies. It also enables a unified interface towards the core network and allows for multi-operator resource sharing. Moreover, the network enables localized communication, one of the critical requirements of rural areas. We finally analyze how the proposed architecture addresses the rural requirements in a qualitative manner.

- In Chapter 3, we present a network architecture to implement the desired network characteristics articulated by Frugal 5G network using the existing Radio Access Technologies (RATs) viz, Third Generation Partnership Protocol (3GPP) 5G network and IEEE 802.11 based Wi-Fi. We refer to this network architecture as *5G-Flow* network. We have analyzed various challenges that arise when we try to realize the Frugal 5G network architecture using 3GPP 5G system. None of the existing multi-RAT network architectures [59-68] address these challenges. Therefore, we realize the Frugal 5G network by complementing the 3GPP 5G access network with OpenFlow based controller and switches. With minimal changes in the 3GPP 5G access network and none in the core network, we are able to realize a unified and integrated multi-access 5G-Flow network. The proposed architecture also allows us to completely decouple communication of User Equipment (UE) with core network from its communication with access network enabling a UE to use any access technology to connect to any core network (say, use 5G New Radio (NR) access to connect to 4G core, which is not possible in the 3GPP architecture) or to directly connect to Internet from access network without going via core. We also discuss additional use-cases of the proposed architecture in the chapter.

- Chapter 4 discusses the evaluation platform that we have developed for analyzing the performance of the 5G-Flow network. We have developed a *system-level simulator for multi-RAT 5G network* in Matrix Laboratory (MATLAB). We have released the source code of the simulator under MIT License and is available online [45]. In our simulator, packets are the fundamental objects, and we have implemented physical and media access layer protocol stacks for 5G NR and Wi-Fi. We have also implemented a centralized controller, as proposed in 5G-Flow network to manage these access technologies. We analyse the performance of data flow management across multiple access technologies in uplink as well as in downlink direction. The results demonstrate that 5G-Flow offers a significant improvement over the existing 3GPP 5G network, realizing an efficient multi-RAT network.

- As the middle mile network is a crucial component of a rural broadband network, we present an algorithm to design the network topology in Chapter 5. We aim to maximize the network throughput along with number of hops and number of middle mile nodes in the network. Minimizing the number of nodes while designing a network topology has exponential complexity. The problem that we intend to solve not only minimizes the number of nodes but also the number of hops while maximizing the network throughput. Therefore, we propose a *simulated annealing based heuristic algorithm* for designing the middle mile network. The performance results reveal that the proposed method outperforms the greedy approach, even in the critical parameters such as minimum user throughput and maximum hop count.

- Frugal 5G network architecture supports multiple operator resource sharing in the network. Towards this, Chapter 6 formulates a spectrum sharing problem among middle mile networks owned by multiple operators. We consider a Long Term Evolution (LTE) based middle mile network operating in sub 1 GHz frequency band, wherein interference is a prominent issue. We present a centralized graph theory based Fairness Constrained Channel Allocation (FCCA) algorithm, employing Carrier Aggregation (CA) and Listen Before Talk (LBT) features. The FCCA algorithm uses a novel concept of allocating a combination of a shared and dedicated channel to an LTE based enhanced NodeB (eNB). We perform extensive system level simulations in network simulator-3 (ns-3) to demonstrate that FCCA not only increases spectral efficiency but also improves system fairness.

We conclude with discussing the directions for future work in Chapter 7. We also discuss P2061, a standards development project initiated by IEEE [46]. P2061 project aims at developing an "Architecture for Low Mobility Energy Efficient Network for Affordable Broadband Access". A significant part of this thesis has been presented as contribution to P2061 standard and is likely to form part of the said standard when it is finalized.

Chapter 2

Frugal 5G Network Architecture

In the previous chapter, we have identified various requirements/challenges pertaining to connectivity in rural areas. Before delving into solutions that address these challenges, it is important to understand the technical requirements (concerning rural areas) identified by the ITU-Radiocommunication sector in the IMT-2020 specifications for 5G networks [47]. Before IMT-2020, i.e., in IMT-2000 and IMT-Advanced, there was no use-case relevant to the rural areas of developing countries. The Rural enhanced Mobile Broadband (eMBB) use-case of IMT-2020, as originally defined, has configurations such as a cell radius of about 1 km and support for high-speed mobility that are inconsistent with the rural connectivity requirements in developing countries. To address this gap, Telecommunications Standards Development Society, India (TSDSI) has proposed Low Mobility Large Cell (LMLC), a new test configuration under the Rural eMBB use case [48]. LMLC test configuration has been added as a mandatory test configuration in IMT-2020 Technical Performance Requirements (TPR). Under LMLC configuration, cell radius has been extended to 3 km, and low mobility (up to 30 km/h) support has been added.

Some of the rural connectivity requirements identified by us, such as low mobility, are aligned with the LMLC test configuration for 5G networks. Moreover, we have identified additional critical requirements for rural areas and aim to address them by developing a novel network architecture. We refer to this network architecture as “Frugal 5G”. The Frugal 5G network architecture presents a conceptual framework that can be realized by utilizing components (with suitable modifications) from the evolving 5G cellular standards and other existing wireless technology standards. We believe that programmable network paradigms can help introduce flexibility and programmability in the network where it is required. Towards this, the Frugal 5G network architecture utilizes Software Defined Networking (SDN) and Network Function Virtualization (NFV) paradigms. It also exploits the cloud and fog architectural model consisting of a cloud-based core network and fog-based access network. The access network is heterogeneous, comprising macro BSs (can be based on 3GPP 5G NR) and WLANs. The backhaul for WLAN APs is enabled via a wireless multi-hop middle mile network. The main contribution of the proposed network is to bring together appropriate technologies (fog, SDN, NFV, wireless middle mile, WLAN, etc.) in a unified framework to address the rural connectivity challenges. Based on the survey of existing literature and available technologies presented in Chapter 1, we understand that the proposed Frugal 5G network architecture is one of the first works to propose an SDN/NFV based network architecture conforming to the rural connectivity needs.

The rest of the chapter is organized as follows. We start by providing a primer on the networking paradigms employed in our work (Section 2.1). We then present an overview of the proposed network architecture (Section 2.2). Further, we describe the main components and working principles of the architecture (Section 2.3 and 2.4). We conclude by presenting a qualitative analysis of our proposal and a summary of this chapter (Section 2.5).

2.1 Background

In this section, we describe state-of-the-art network paradigms which lay the groundwork to better understand the proposed network architecture.

2.1.1 Software Defined Networking

In the traditional communication networks, the control plane (which takes traffic handling decisions) is essentially distributed and coupled with data forwarding functions. Additionally, the control plane is embedded and tightly coupled with the hardware. Therefore, network management becomes complex, and the introduction of new functionalities

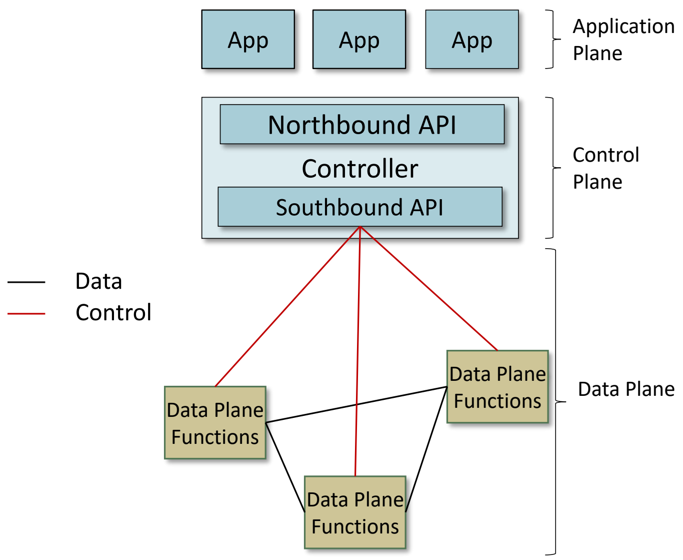


Figure 2.1: Overview of a software defined network

in the network is very challenging and time-consuming. To remedy this, the idea of programmable networks has emerged to make the network more flexible. SDN, a programmable network paradigm, separates the *control plane* functions from the *data plane* functions to simplify the network management. A graphical depiction of a software-defined network is shown in Figure 2.1. Under the SDN paradigm, a logically centralized controller is introduced and is responsible for decision making. The data plane, comprising forwarding network devices from which control plane functionalities have been removed, is only responsible for data forwarding based on the rules set by the controller. An open and standardized interface separates the control plane and data plane and is called *Southbound Application Programming Interface (API)* as shown in Figure 2.1. Southbound API provides the instruction set used to program the forwarding devices. OpenFlow is an established Southbound API. The *Northbound APIs* presents an abstraction of underlying network resources and network topology to the application plane. APIs such as Rest, Python, Java are commonly used Northbound APIs.

Now, we list the major benefits of the SDN paradigm. The abstraction of the control plane provided by SDN simplifies the modification of network policies via high-level languages. Additionally, more effective and sophisticated policies can be designed as the SDN controller provides a global view of the network. SDN also enables easier integration of various policies in the network, such as routing and load balancing. In addition to these benefits, SDN allows for network management at a granular level. However, the most important benefit of SDN, in the context of rural areas, is reduced network cost. Due to the abstraction of the control plane, the data plane can be implemented using low-cost switches with limited processing capabilities.

2.1.2 Network Function Virtualization

Conventionally, the network functions such as firewalls, Quality of Experience (QoE) monitor, Provider Edge (PE) router, network address translator, load balancers are built on proprietary hardware customized for these network functions. As a result, the network functions can not share the hardware resources, leading to resource wastage. Additionally, designing and deploying new network functions is challenging, expensive, and time-intensive. NFV, another programmable network paradigm, can address these challenges by abstracting network functions from proprietary hardware platforms and implementing them in software. This abstraction allows network operators to design network functions that can run on commercial off-the-shelf equipment.

European Telecommunications Standards Institute (ETSI), a European standards organization, has defined an NFV framework, and it is illustrated in Figure 2.2 [49]. NFV framework comprises Network Function Virtualization Infrastructure (NFVI), Virtual Network Functions (VNFs) and NFV Management and Orchestration (MANO). NFVI consists of software as well as hardware resources that provide an environment for instantiation of VNFs. Virtual resources are abstracted from the physical compute, storage,

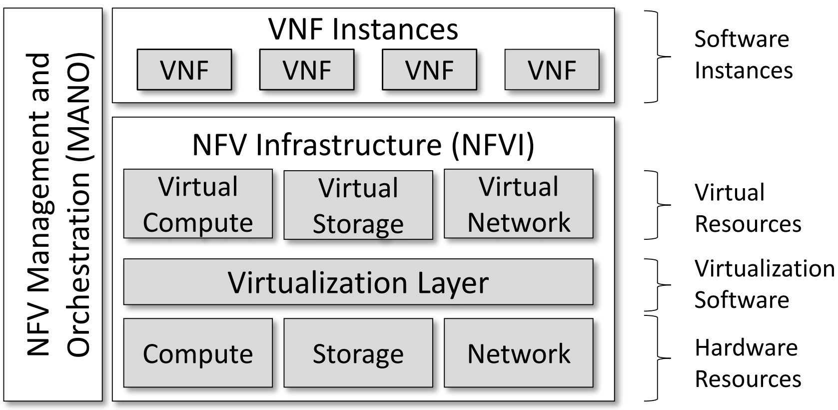


Figure 2.2: Overview of NFV framework

and network resources with the help of virtualization layer. VNF is an implementation of a network function that is instantiated on these virtual resources. NFV MANO is the most complex component of the NFV framework. It is responsible for the onboarding and lifecycle management of network services. It also enables lifecycle management of VNF instances. Besides, it manages the compute, storage, and network resources.

NFV offers many advantages, and they are described as follows. Virtualization of the network resources enables flexibility in increasing and decreasing these resources as required by a network function. With the use of commercial off-the-shelf hardware, capital expenditure can be significantly reduced. Moreover, the hardware can be added or reduced based on the requirement, making the system cost-effective. As multiple network functions can run on a single hardware, the power consumption and space utilization is also reduced. Application of NFV enables faster innovation cycle and quick roll-out of services. It also enables sharing of resources among multiple users, applications, and operators.

2.1.3 Fog Computing

To understand fog computing, it is essential to first understand cloud computing. Cloud computing enables users/applications to access computing, storage, and network resources on demand. Cloud computing has several benefits, including cost-effectiveness, flexibility, and elasticity in using the available resources. However, the exponential growth of data in the past few years has demanded the extension of computing, storage, and network resources closer to where the data is being generated. As a result, distributed applications with requirements such as low latency and high bandwidth can be supported. Fog computing addresses this gap by introducing computing paradigms closer to the end-user devices. OpenFog Consortium defines fog computing as "a horizontal system-level architecture that distributes computing, storage, control, and networking functions closer to the users along a cloud-to-thing continuum" [50].

Due to the local data handling at the fog node, a considerable amount of backhaul bandwidth can be saved, and low-latency applications can be easily supported. The applications running on a fog node can be designed and deployed where they are needed (based on the requirements of the end-users).

2.2 Frugal 5G Architecture Overview

We first explain the Frugal 5G network architecture with the help of a deployment example given in Figure 2.3. In our proposed network architecture, we mainly focus on the access network and describe the various elements of the access network. We consider a heterogeneous access network comprising macro BS and several WLANs. We assume that a PoP is available in the vicinity of the given rural area. A macro BS is co-located with the PoP and provides blanket coverage to the end-users in a large geographic area. It also provides the required Quality of Service (QoS) for critical applications. Owing to the clustered settlement of people in rural areas, WLAN is deployed only in clusters. IEEE 802.11, the most commonly used family of WLAN standard, can be chosen for the last mile as it facilitates fixed high-speed broadband access to the end-users cost-effectively. The major challenge in designing an access network is to backhaul the traffic generated by WLAN APs. This is enabled via a wireless middle mile network that connects WLAN APs to a PoP. A middle mile AP is located near a PoP that wirelessly communicates with middle mile clients (located in the village clusters), which in turn serve WLAN APs. The middle mile network supports optimized multi-hop mesh network topology, addressing the problem of backhauling WLAN APs effectively. WLAN and middle mile APs/clients are low power nodes and can cost-effectively run on renewable energy sources. The architecture proposed here (Figure 2.3) is based on the rural settlement pattern[1] shown in Figure 1.5 and helps in the efficient delivery of services to the users.

A significant part of the control and management of this heterogeneous access network is performed closer to the end-user by the fog node. The deployment of intelligence at the edge via fog node enables more efficient communication with reduced resource usage, such as routing end-to-end data flows locally within the access network wherever applicable. Cache servers can be instantiated in fog, enabling the cost-effective delivery of relevant regional content from cache servers. Moreover, fog can enable local communication even when connectivity with core or external data network is unavailable.

[1]Since we aim to illustrate a sketch for the deployment scenario, we assume arbitrary number of village clusters in the example. The actual number of village clusters served by the network depends on the various factors such as network technology and rural demography.

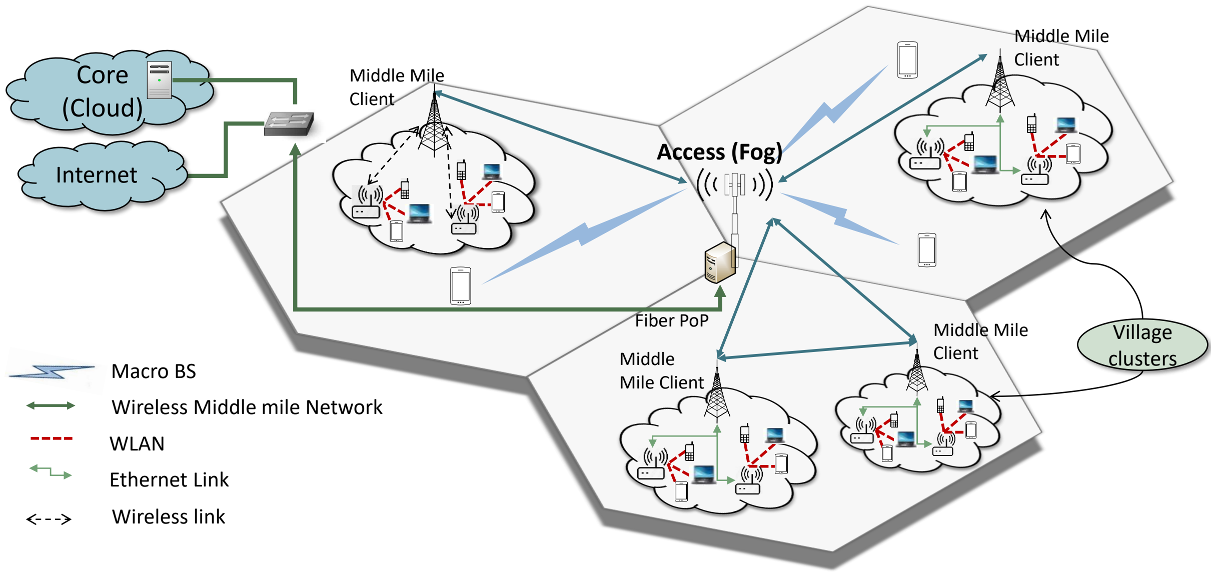


Figure 2.3: A deployment example of Frugal 5G network.

We have employed network paradigms viz., SDN, and NFV to design the fog node. SDN separates the control and data plane functionality in the fog. Following this separation, we can unify the control functionalities of the proposed heterogeneous network. This unified control introduces efficiency in the delivery of services. Depending on the service requirements and the network condition, the individual data flows are dynamically setup through appropriate data plane elements in the access network. Moreover, the decoupling of control and data plane facilitates independent evolution and development of control/-data plane entities. A unified SDN controller may also enable energy saving by driving a subset of WLAN APs and middle mile clients into power save mode. All the control and data plane functions in the fog are instantiated as virtual network functions that can run on commercial off-the-shelf platforms. Typically, a cellular network is upgraded every few years to keep up with the technological advances. If virtualization is enabled, these upgrades will happen mostly in software with minimal hardware upgrades, keeping the physical infrastructure intact, consequently reducing the capital expenditure. This can also aid operators in enabling regional services that would, otherwise, have been difficult to implement on dedicated hardware.

Figure 2.4 provides an essence of the proposed network architecture and its elements. The proposed architecture possesses a hierarchical structure comprising a centralized cloud element and multiple distributed fog elements located near the users. Both the fog and

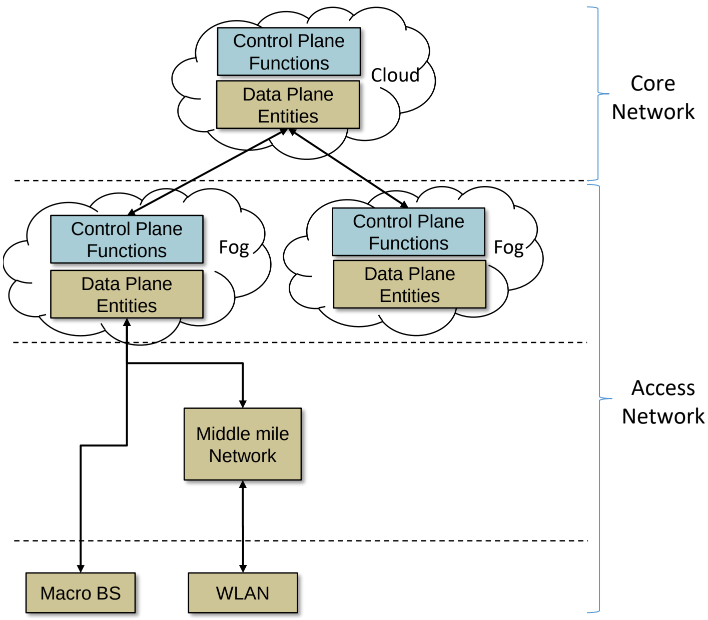


Figure 2.4: Frugal 5G architecture concept

the cloud use SDN-based architecture with well-defined separation between control and data plane entities. The centralized cloud hosts the core network and performs a set of functions, including the ones performed by traditional wireless core networks. A fog comprises access network elements belonging to different RATs. A controller in each network element controls and manages data plane entities from multiple RATs.

2.3 Main Components

We now describe the primary building blocks of Frugal 5G network architecture as shown in Figure 2.5. We categorize all network functions in the access network and core network under control plane or data plane functions. The core network is responsible for the overall control of Frugal 5G network. The control plane of the core network performs standard tasks, such as mobility management, policy, and charging control, similar to the core network of other cellular systems. The core network sets up the traffic flow for a user through these data plane entities. It can also set up a path between two network elements if required. Additionally, the core network control plane manages the distributed SDN controllers in each of the fog nodes. The data plane functions in the core network are responsible for forwarding the received data to another data plane entity in the network or the external data network. The control and data plane functions across the network

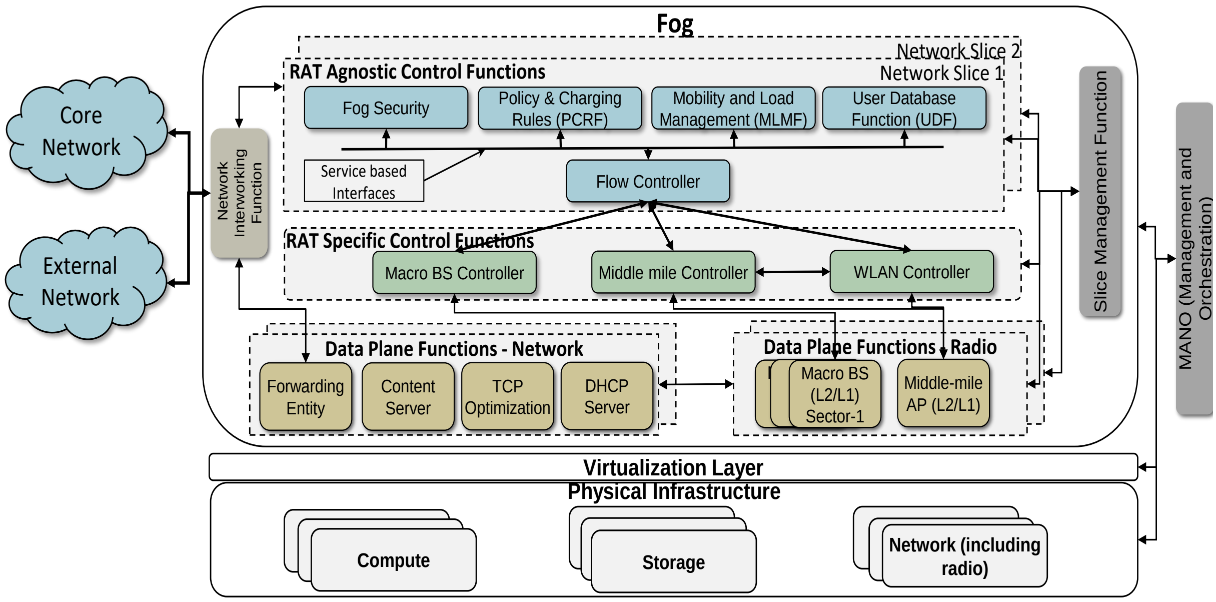


Figure 2.5: Detailed architecture of access network (fog)

and cloud are synchronized to avoid any inconsistencies. We now describe, in detail, the working of control and data plane in the access network.

2.3.1 Control Plane of the Access Network

As shown in Figure 2.5, we propose a layered SDN controller in which the top-level functions have a unified view of the multi-RAT Radio Access Network (RAN) and are responsible for network-wide decision making. We refer to these functions as *RAT Agnostic Control Functions (RACFs)*. The decisions taken by RACFs are translated into RAT specific decisions by the lower level *RAT Specific Control Functions (RSCFs)* and supplied to the corresponding data plane functions. The proposed architecture also supports a *unified interworking function* towards core/external data network. We now discuss in detail the various components of the control plane.

RAT Specific Control Functions

RSCFs provide an abstract view of the underlying access network to the higher-level control entities such as RACFs and the Slice Management Function. The abstract view of the resources enables a unified control of the multi-RAT by RACFs. It also enables virtualization of the RAN and facilitates the creation of multiple logical networks or network slices. RSCFs perform the tasks such as device configuration, information broadcast, resource allocation, and QoS management. It is important to note that these control functions primarily execute the decisions taken by the higher level RACFs.

RAT Agnostic Control Functions

The flow controller is an RACF that directly interfaces with RSCFs. It is responsible for allocating resources and path establishment for individual traffic flows through the network. It operates over the abstract resources provided by the lower level functions, i.e., RSCFs. The flow controller analyses the individual traffic flows and associates QoS attributes with each of them with the help of Policy and Charging Rules Function (PCRF) and User Database Function (UDF). For instance, a traffic flow may be categorized as a real-time flow along with a stipulated guaranteed data rate. This flow-related information may be provided to other RACFs such as Mobility and Load Management Function (MLMF), which may utilize the information to perform the mobility and load management tasks. MLMF has a global view of all connected users, which includes their location and the RAT association. UDF maintains the user profile for each of the subscribers, which is accessed to perform functions such as user authentication and QoS determination.

In traditional cellular networks, PCRF, UDF, and part of MLMF are located in the core network. However, to enable localized communication under individual fog elements, these functions can be instantiated (in fog) if the resources to house them are available. The RACF instances in the fog, like PCRF and UDF, are synchronized with their counterparts in the core network (in the cloud) regularly to avoid any inconsistency in control information kept across the fog and cloud. Since sensitive user data has been moved from core to access network (edge of the network), we need to employ a network security function responsible for secure access to the sensitive user data in fog.

Interfaces

There are three types of interfaces in the proposed network.

1. The interactions between RACFs are enabled via service based interfaces and may use RESTful APIs [51] to get the required information from each other.

2. The interface between RSCFs, e.g., macro BS controller, or WLAN controller and the corresponding data plane entities may be analogous to F1-Application Protocol

(F1-AP) (a 3GPP defined interface between 5G Next-generation NodeB (gNB)-Centralized Unit and gNB-Distributed Unit) [52] or Control And Provisioning of Wireless Access Points (CAPWAP) protocol.

3. The interface between flow controller and RSCFs is analogous to OpenFlow interface. OpenFlow protocol [51] may need some modifications to be applicable in this scenario.

Slice Management Function

The Slice Management Function creates multiple logical networks or network slices utilizing the abstract resource view provided by RSCFs. The network slices may be formed across multiple operators or it may be service-based for a single operator. In rural areas, sharing of network infrastructure among multiple operators can considerably lower the network cost. The method of slicing the network resources is discussed, in detail, in Section 2.4.4.

2.3.2 Data Plane of the Access Network

The data plane in Frugal 5G network broadly includes data forwarding and service functions, such as Dynamic Host Configuration Protocol (DHCP) server. The functions instantiated in the network include forwarding elements of macro BS and middle mile AP. The forwarding elements of WLAN APs and middle mile clients are located in the village clusters. The forwarding elements of different RATs in the access network are controlled by their respective control functions instantiated in the network. Various service functions are part of the network's data plane, such as DHCP server, content (cache) server, and TCP optimization functions. In traditional cellular networks, DHCP server is located in the core network because core network is responsible for the session management and IP allocation. To enable local data flow without involving core network, DHCP server is also instantiated in the network. User experience may improve significantly by keeping these service functions near the users (in the distributed network elements) as they reduce the end-to-end latency. This also reduces the backhaul usage, thereby decreasing the operational cost of the network.

2.4 Key Working Principles

As shown in Figure 2.6, there are five key working principles of Frugal 5G network which makes it a suitable solution for a rural broadband network. These working principles are discussed in detail next.

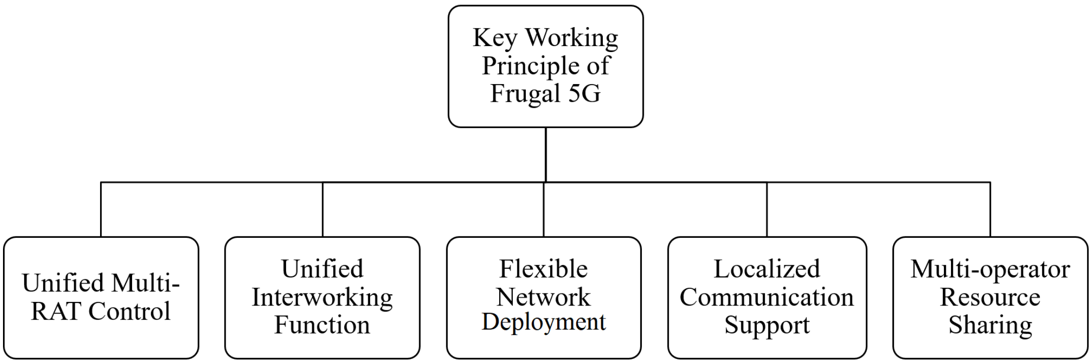


Figure 2.6: Key working principles of Frugal 5G network

2.4.1 Unified Multi-RAT Control

The proposed Frugal 5G network comprises multiple RATs, i.e., macro BS, WLAN, and middle mile network. The proposed architecture integrates multiple RATs under a unified framework, wherein radio resources provided by these RATs can be optimally utilized. A unified control may help in additional network optimization tasks, e.g., energy saving, by driving a subset of WLAN and middle mile nodes in sleep mode during low traffic scenarios. Unlike relays in cellular networks that support no more than two hops, the middle mile network is a multi-hop mesh network. Therefore, integrating the middle mile network with the other RATs in our proposal is crucial in facilitating the required QoS. When a data packet has to be delivered to a user through WLAN, the flow controller gives the required instructions to both WLAN and middle mile controllers, and an appropriate data path is set up through the middle mile network and WLAN. Note that, middle mile network provides an infrastructure for WLAN APs to communicate with a WLAN controller as well. The control information exchange between a WLAN controller and WLAN APs takes place over the same middle mile network, which also carries the data for users.

2.4.2 Unified Interworking Function

As shown in Figure 2.5, there is a unified interworking function that interfaces with the core network, simplifying the interworking with the core network. Additionally, a unified interface is crucial for realizing a unified control of the multi-RAT access network. The access network under the fog node may be seen as a single entity in the larger network for the core network. The decisions taken by access network are transparent to the core network and hence, greater flexibility is enabled at the access network level to manage a multi-RAT network. With a unified interworking function, a user under Frugal 5G access network may directly communicate with the external data network, bypassing the core network.

2.4.3 Flexible Network Deployment

The usage of NFV in the access network (fog) and core network (cloud) allows for a flexible instantiation of functions across fog and cloud. The SDN controller in the cloud provides policy configuration to the management and orchestration entity to instantiate different network functions on the individual fog elements. The access network architecture, as shown in Figure 2.5 is one of the possible ways in which network functions can be instantiated. Their instantiation in the network may vary and is dependent on the availability of the resources. For instance, when the compute and storage resources are limited, only a handful of network functions, e.g., radio specific data plane functions, forwarding functions, core network interworking function, etc., are instantiated in a fog element. On the other hand, if the compute and storage resources are available but the backhaul bandwidth (network resources) is limited, several functions can be installed in the fog, particularly the functions that help reduce the backhaul usage, e.g., the content server.

2.4.4 Multi-operator Resource Sharing

Sharing network resources is an important technique that can considerably reduce the network cost. The proposed Frugal 5G network architecture allows us to split the physical network resources into multiple logical networks, also called network slices. Each of these network slices or logical networks can be used for different purposes, e.g., the

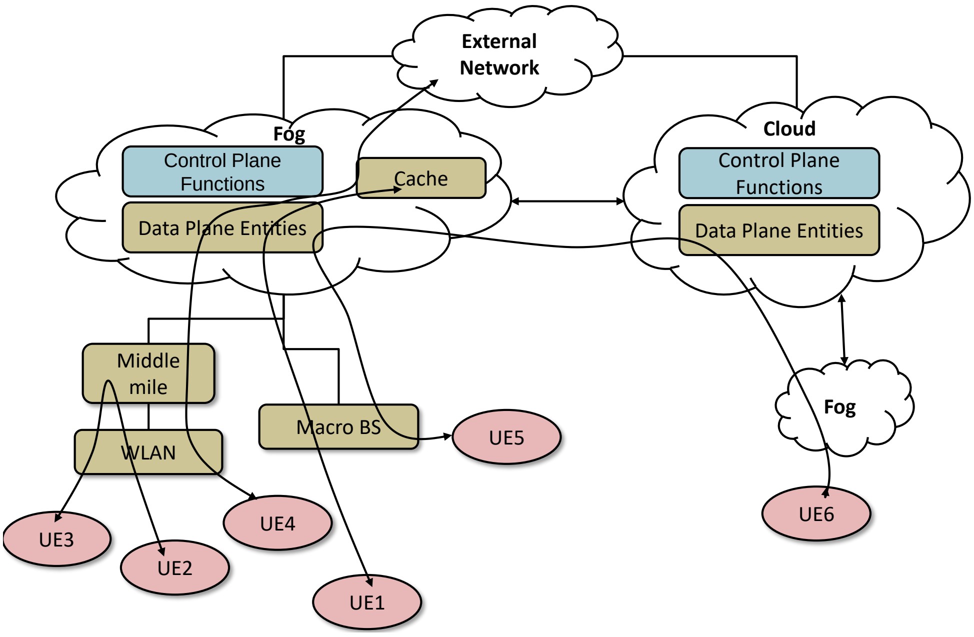


Figure 2.7: Dataflow examples

individual slices can be allocated to an operator through which the operator can serve the rural areas. It is crucial to understand how the network resources are split among multiple slices in our network. Depending on the various parameters, such as subscriber base, an operator might need a set of network resources. As already mentioned, RSCFs provide an abstract (virtual) view of the underlying network resources to the slice management function. The slice management function splits these virtual network resources into multiple network slices, each of which is allocated to individual operators. RACFs are instantiated separately for each of these slices as they may be specific to the operator's network. This method of network slicing provides complete flexibility in terms of network resource sharing. When one operator's network (slice) experiences a low load scenario, its resources can be shared with the other operator's network (slice), which may be experiencing a higher load. Since the proposed network architecture uses more than one RAT, multiple network slicing techniques may be employed. For example, WLAN can be sliced based on Service Set Identifiers (SSIDs)/Virtual Access Points (VAPs). Macro BS can be sliced based on radio resources.

2.4.5 Localized Communication Support

In remote areas, it may take a significant time to resolve the connectivity issues such as backhaul failures. To enable localized communication in such a situation without the involvement of core network, we have instantiated relevant network functions in the fog, as explained in Section IV. This results in improved backhaul efficiency (by avoiding the cloud-fog signaling/data transfer) and user experience. To understand how the localized data paths are handled by fog, it is essential to understand the data/traffic flow set-up in the network. Various possible data/traffic flows in the network are illustrated in Figure 2.7. SDN controllers are responsible for setting up the data path through the network for a user-specific traffic flow. Depending on various parameters, such as user's mobility, strength of the radio signal, QoS requirements, and load on different RATs, the flow controller selects an appropriate data path for a user in the access network. A data path for a user is selected either through macro BS or through wireless middle mile network and WLAN AP. Now, it is possible that an end-to-end data path for a traffic flow is fully contained within a single network element. An example for such paths may be a Voice over IP (VoIP) session between two mobile users, who are in the close vicinity of each other (like UE2 and UE3 in Figure 2.7) or a data session between a user and the content (cache) server (like UE1 in Figure 2.7). Usage of such data paths reduces the end-to-end latency of the flow and optimizes the resource utilization in the network by not using the wireless backhaul and the core cloud elements. Since network elements can continue to provide services of localized nature in the absence of connectivity to the cloud, network resilience is improved. Thus, we may reduce the cost of network equipment by employing limited fault tolerance in the equipment.

2.5 Analysis of Frugal 5G Network

2.5.1 Qualitative Analysis

Since Frugal 5G is an abstract network architecture providing design guidelines for a rural broadband network, we analyze its performance in a qualitative manner. Table 2.1 lists the various connectivity requirements of rural areas (discussed in Chapter 1) and the supporting features in the Frugal 5G network architecture that enable these requirements.

Table 2.1: Summary of how Frugal 5G network architecture addresses the rural requirements

Requirements	**Supporting Features in Frugal 5G Network Architecture**
Affordability	• Low-cost low-power WLAN and middle mile nodes • Multi-operator resource sharing reduces capital expenditure • Macro BS can operate on lower frequency sub-Gig band to achieve large coverage area, with low power transmission • SDN/NFV enables usage of inexpensive commodity hardware
Support for clustered service provision	• High-speed WLAN connectivity only in clusters • Few macro BSs provide carpet coverage
Support for regional content generation and storage	Cache enabled network element allows for local content generation, storage and access
Limited mobility	• High-speed connectivity via WLAN backhauled by fixed wireless middle mile network • Macro BS provides limited mobility support resulting in a less complex system [43]
Connectivity to remote/isolated areas	Wireless middle mile network allows connectivity to be established in remote/isolated areas where it is difficult to provide wired connectivity
Localized communication	Multiple RATs i.e. WLAN, wireless middle mile network and macro BS are controlled by a unified SDN based control framework in network enabling efficient routing of traffic flows through the access network and also allowing for the set up of localized communication paths, wherever possible/needed.
Capable of running on renewable energy	Low-power middle mile and WLAN nodes can operate on renewable energy in a cost-effective way

2.5.2 Comparison of System Level Costs

In this section, we provide a rough cost estimate for the Frugal 5G network to understand the order of cost savings. Further, we compare the cost estimates of the main network elements of the Frugal 5G network with that of the network elements in a standard cellular network (4G/5G). For this cost analysis, we consider the case study of the Palghar testbed. The rural area considered in the Palghar testbed spans 25 sq. km and consists of seven villages (Section 1.4). We assume an optical PoP (centrally located) that can be used to provide connectivity in the given area.

Frugal 5G Network: To provide a carpet coverage (for basic voice and data connectivity) in 25 sq. km area, we consider a macro BS of 3 km radius. Coverage of 3 km can be easily supported if the macro BS operates in sub-1 GHz frequency bands such as the TV UHF band. To provide high-speed connectivity in village clusters, we employ Wi-Fi APs. Let us assume each village cluster might require four Wi-Fi APs on average. Therefore, a total of 28 Wi-Fi APs are required to serve seven villages in the given area. As proposed in the Frugal 5G network architecture, a wireless middle mile network backhauls Wi-Fi APs. Specifically, a middle mile client is deployed per cluster that backhauls all the Wi-Fi APs in that cluster via wired connectivity. We assume an IEEE 802.11 based multi-hop middle mile network incorporating approximately ten middle mile nodes.

Standard Cellular Network: To provide coverage in a given rural area and enable the required capacity, cellular operators would typically deploy multiple macro BSs. Note that a single macro BS considered in the Frugal 5G network may not provide the required capacity. The Inter-Site Distance (ISD) of a rural macro BS is 1732 m [48]. To approximately cover the area of 25 sq. km, we would require seven macro BSs assuming a 2-tier deployment. There is also an additional cost of laying the optical fiber to backhaul these macro BSs. The central macro BS (in first-tier) is co-located with the optical PoP, but a macro BS in the second-tier is approximately 1 km away.

Table 2.2 compares the total cost of network elements for above-mentioned scenarios. As shown in the table, the cost estimate for the network elements in a standard cellular network is approximately seven times that of the Frugal 5G network. This provides a good insight into the affordability of the proposed Frugal 5G network architecture. Further, if

Table 2.2: Comparison of System Level Costs

Network Element	**Cost per unit (USD)**	**Frugal 5G Network**		**Standard Cellular Network**	
		Quantity	**Cost (USD)**	**Quantity**	**Cost (USD)**
Macro BS	50,000 [53]	1 unit	50,000	7 units	350,000
Wi-Fi APs (outdoor)	150 [54]	28 units	4,200	-	-
Middle mile network[2]	200 [55]	10 units	2,000	-	-
Optical Fiber[3]	7080 (per km) [56]	-	-	6 km	42,480
Total Cost			**56,200**		**392,480**

we consider the deployment, operation, and maintenance costs, the overall cost-savings delivered by the Frugal 5G network may be even higher. A comprehensive economic analysis examining the various costs mentioned above for multiple deployment scenarios, use-cases, and data rate requirements is not the scope of this thesis. However, we aim to address this in the future.

2.6 Summary

Based on the connectivity challenges/requirements of rural areas, we have proposed an innovative communication network architecture referred to as the Frugal 5G network in this chapter. The proposed network architecture addresses the concerns of rural connectivity through its unique and novel features—unified control framework with SDN, wireless backhaul RAN integration, the capability of isolated operation and localized communi-

[3]We have not included the cost of wired connectivity between middle mile client and Wi-Fi AP as it is quite inexpensive.

[3]The cost of laying optical fiber includes cost of digging, insulation pipes, cost of fiber and right of way.

cation, flexibility in service/function instantiation in the network, and the framework for network virtualization and network slice creation.

Chapter 3

5G-Flow: Realizing the Frugal 5G Network

In the previous chapter, an abstract network architecture named Frugal 5G has been presented that specifies the desired network characteristics to address the rural connectivity challenges. In this chapter, we discuss the realization of the Frugal 5G network using existing technology standards to analyze its feasibility. Since 3GPP 5G network is an upcoming cellular network and is likely to be widely deployed, we consider 5G NR based gNB as macro BS. To enable high-speed connectivity in clusters, we employ IEEE 802.11 based Wi-Fi APs. To realize the desired characteristics of Frugal 5G network architecture (discussed in Chapter 2), it is essential to understand the implementation challenges. Frugal 5G network features a unified multi-RAT network at the access network. Although 3GPP 5G network integrates multiple RATs comprising 3GPP based access (e.g., gNB and eNB) and non-3GPP based access (e.g., Wi-Fi and wireline), these RATs are integrated at the core network level. Moreover, there is no support for a unified interworking function towards the core network, an important characteristic of the Frugal 5G network. Therefore, several challenges need to be addressed to successfully realize the Frugal 5G network within 3GPP 5G framework. The existing literature that proposes a network architecture for multi-RAT integration is either limited to IEEE 802 family (not extendable to cellular networks) [57–59] or present an entirely new framework along with changes to radio protocol stack and hence suffer from implementation challenges [60–64].

On the other hand, we intend to propose a simpler implementation of the Frugal 5G network architecture by proposing minimal software-based changes in 3GPP 5G network.

To realize the Frugal 5G network, we supplement the existing 3GPP 5G architecture with OpenFlow, an established southbound interface for SDN-based networks. We refer to the proposed network architecture as *5G-Flow*. It employs OpenFlow switches (network switches based on OpenFlow protocol) and a light-weight OpenFlow controller (also called 5G-Flow controller). The flow controller (the principal *RAT agnostic control function*), responsible for system-wide decision making in Frugal 5G network, is implemented via the 5G-Flow controller. To simplify the implementation of *RAT specific control functions*, which execute the decisions taken by the 5G-Flow controller, we use the existing radio protocol stack of gNB and Wi-Fi. The OpenFlow switch introduced at the network side integrates multiple RATs and realizes a unified interworking interface towards the 5G Core Network (5GC). We have also introduced an OpenFlow switch at each UE, wherein UE's connectivity to RAN is fully decoupled from its connectivity to the core network. This brings immense flexibility and enables a UE to interface with 5GC, 4G core network, Internet, or any other data network via any 4G/5G/Wi-Fi based RAN. A light-weight 5G-Flow controller enables RAN level management of downlink as well as uplink dataflows, in turn, utilizing the multi-RAT resources efficiently.

The rest of the chapter is organized as follows. We begin by discussing the existing literature that deals with applying SDN concepts for designing a multi-RAT network architecture (Section 3.1). We briefly discuss OpenFlow protocol and some important 3GPP 5G terminologies (Section 3.2). We then discuss the implementation challenges that surface when Frugal 5G network is realized within 3GPP 5G framework (Section 3.3). We discuss the proposed network architecture and its working in detail (Section 3.4 and 3.5). We discuss additional use-cases of the 5G-Flow network next (Section 3.6). We conclude by presenting the summary of the chapter (Section 3.7).

3.1 Prior Work

Several efforts in literature deal with the application of SDN in cellular networks to make the network more programmable and in turn, more manageable. Research works such as SoftRAN [65], SoftMobile [66], FlexRAN [67], SoftNet [68] and Software-Defined Wireless

Network (SDWN) [69] aim at making RAN more programmable, flexible and specifically, resources efficient. SoftAir [70] uses SDN paradigm and hardware abstraction to enable a scalable network architecture for 5G. OpenFlow is used to enable software-defined base stations and switches in the network. SoftCell [71] aims to introduce scalability and flexibility in cellular core networks. To support rapid protocol evolution, OpenRadio uses SDN paradigm to decouple the wireless protocol into processing and decision plane, and provide a programmable interface between two planes [72]. In [73], authors apply the concept of flow-based forwarding model in cellular networks. To manage IEEE 802.11 based networks, Odin network architecture has been proposed in [74]. It introduces Light Virtual Access Point (LVAP) abstraction for seamless mobility of users. None of the above works apply SDN for multi-RAT integration in the access network.

We now discuss the research works that propose an SDN based network architecture to manage heterogeneous RATs. In [57], authors introduce a virtual MAC layer at network nodes and users in order to manage heterogeneous RATs in a technology independent manner. In [58, 59], the authors have developed a prototype that augments WiMAX and Wi-Fi APs with OpenFlow. Mobility management of heterogeneous RATs has been evaluated as an application. However, the extension of the proposed concept for 4G/5G network is not straightforward. OpenRAN [60] uses virtualization to design a software-defined multi-access RAN. The authors in [61] propose an end-to-end SDN based architecture for a mobile network with a controller managing both access and core network data plane elements by providing an abstract view of the underlying network resources to the controller. 5G-EmPower [62] proposes an SDN based framework to manage heterogeneous access networks by abstracting the technology-dependent features of RATs. In our work, we suggest minimal changes to 5G RAN while 5GC remains unaffected. Also, we do not intend to propose a new framework but suggest enhancements to the existing 3GPP 5G architecture to realise a multi-RAT RAN.

In [63], the authors propose a multi-RAT architecture and define a unified and open interface between the controller and multi-RAT data plane. In [64], authors propose a convergence sub-layer over layer 2 of multiple RATs in order to tightly integrate them. To integrate LTE with Wi-Fi, LTE-Wireless Local Area Network Aggregation (LWA) mechanism has been proposed by 3GPP [75, 76]. LWA proposes an emulation layer over Wi-Fi

AP that encapsulates LTE packets in Wi-Fi Media Access Control (MAC) frame. All of the above proposals [63, 64, 75, 76] modify the radio protocol stack for RAT integration. In our work, the radio stack of different RATs in 5G-Flow RAN remains unchanged.

None of the above proposals suggest enhancements to the recently defined 3GPP 5G RAN architecture to the best of our knowledge. Moreover, use of OpenFlow to propose a unified interworking entity and thereby integrating multiple RATs (with no changes in radio stack) in downlink and uplink has also not been discussed in literature. Decoupling UE's communication with RAN from its communication with core network within 3GPP 5G framework is also one of the novel contributions of our work.

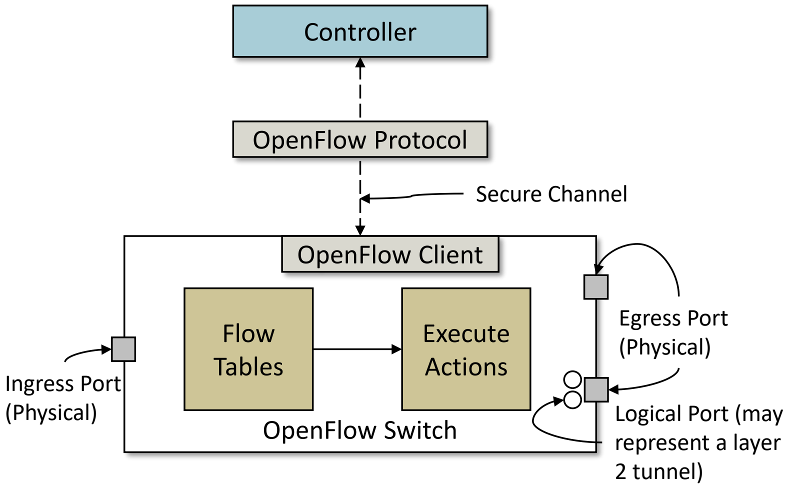
3.2 Background

In this section, we discuss OpenFlow protocol and some basic 3GPP 5G terminologies which form the basis of our discussion ahead. We also discuss the implementation challenges concerning 3GPP 5G network.

3.2.1 OpenFlow Protocol

OpenFlow protocol is an open interface that allows an SDN controller to manage the forwarding plane of a network [77]. An OpenFlow compliant network consists of a logically centralized OpenFlow controller and multiple OpenFlow switches. An OpenFlow switch supports an OpenFlow client, which communicates with the controller through a secure channel as shown in Figure 3.1. The OpenFlow switch consists of flow tables which are managed by the controller via OpenFlow protocol. A flow-table comprises several flow-entries that match a packet based on match-fields such as IP address or TCP port number. The controller can add, delete or update the flow entries. Based on the flow-entry, an action is taken (e.g., forward or drop) on the matched packet.

An OpenFlow switch supports physical and logical ports as depicted in the figure. A physical port corresponds to a hardware interface on an OpenFlow switch from where packets can enter or exit. A logical port does not directly correspond to a hardware interface, but it is an abstraction of a physical port. It can be used to implement the processing of packets in OpenFlow switch. *In our proposal, the complex protocol processing is done as part of the logical port in the OpenFlow switch.*



(a) Components of an OpenFlow switch

Message	Message Type	Description
Hello	Symmetric	Message exchange upon connection startup
Packet-in	Asynchronous	Transfer the control of packet to controller
Packet-out	Controller to Switch	Send packet out of specified port on a switch and forward packet received through Packet-in
FlowMod	Controller to Switch	Used to manage flow tables
Port Status	Asynchronous	Inform the controller about change of port

(b) Few examples of OpenFlow messages

Figure 3.1: Introduction to OpenFlow protocol

The OpenFlow protocol supports three types of messages – *controller to switch*, *asynchronous* and *symmetric* messages. Controller to switch messages are used to manage the flow tables in the switch. These messages are initiated by the controller. Asynchronous messages are initiated by the switches and are used to inform the OpenFlow controller about an event. Symmetric messages are used for detecting any problem between OpenFlow controller and switches. Some example OpenFlow messages are described in the table in Figure 3.1.

OpenFlow-Configuration (OF-Config) protocol, a complementary protocol based on Network Configuration Protocol (NETCONF) [78], helps in configuration of OpenFlow switches [79]. OF-Config is responsible for the association between a controller and an OpenFlow switch and the configuration of physical and logical ports. OF-Config can configure virtual OpenFlow switches on an OpenFlow capable switch, where each virtual switch may be controlled by the same or different controllers. To inform the OpenFlow controller about various events at OpenFlow switches such as link failures or configuration changes, OpenFlow Notifications Framework provides a method to subscribe to asynchronous messages based on predefined events [80].

3.2.2 3GPP 5G Multi-RAT Network Architecture

The 5G cellular standard envisions to support various use-cases, i.e., enhanced mobile broadband, ultra-reliable low latency communications, and massive machine-type communications [32]. These new requirements have led to several innovations and a redesign of the cellular network as part of 3GPP 5G standardization, such as service-based architecture, virtualized network functions, control and user planes separation, and network slicing. One of the major advancements is the unification of multiple access technologies in 5GC, which is an essential step towards enhancing network efficiency. The 3GPP 5G access network incorporates heterogeneous access technologies, i.e., 3GPP access (e.g., gNB, and LTE based eNB) and non-3GPP access (e.g., Wi-Fi and wireline access). Towards this, a common interface between 5GC and 5G RAN has been introduced to integrate multiple access networks at the core network as per 3GPP 5G specifications.

We now describe RAN components (both 3GPP and non-3GPP access) in a 5G network along with their protocol stacks and functioning. Figure 3.2 illustrates the protocol stack of UE and gNB as defined in the 3GPP 5G specifications [32]. gNB, a new node providing NR data plane and control plane protocol terminations towards the UE, consists of a Centralized Unit (CU) and one or many Distributed Units (DUs). gNB-CU can be further divided into control plane (gNB-CU-CP) and user plane (gNB-CU-UP) components. UE and gNB both have Radio Resource Control (RRC) layer that facilitates control plane interaction between the two entities. RRC also helps in radio connection establishment and release, mobility management, and setting up of radio bearers. Service

Data Adaptation Protocol (SDAP) layer, along with the underlying protocol stack, at UE and gNB, is responsible for user plane data transfer over the radio interface along with QoS handling of dataflows. Non-Access Stratum (NAS) layer present at UE is responsible for non-radio related signaling between UE and 5GC.

5G RAN communicates with 5GC through N2 and N3 interfaces. Next-Generation Application Protocol (NGAP) layer, together with the underlying protocol stack (N2 interface), is responsible for all signaling (control) message exchange between RAN and Access and Mobility Management Function (AMF). AMF is responsible for UE's mobility management, registration and provide NAS message termination for UE. NGAP layer uses Stream Control Transmission Protocol (SCTP) for transport signaling messages over IP network. Data packets between RAN and User Plane Function (UPF) (the data plane function in 5GC) are exchanged over N3 interface using GPRS Tunneling Protocol (GTP) and the underlying UDP/IP protocol stack.

For non-3GPP access, RAT-specific interworking entities have been defined to interface with 5GC (N2/N3 interface). For instance, a network entity called Non-3GPP Interworking Function (N3IWF) has been introduced for untrusted non-3GPP access (e.g., Wi-Fi). Similarly, trusted Wi-Fi access uses Trusted Non-3GPP Gateway Function (TNGF), and wireline access uses Wireline-Access Gateway Function (W-AGF) to interface with 5GC. It is important to note that interworking entities such as N3IWF incorporate N2/N3 protocol stack, while for 5G NR based access, N2/N3 interface protocol stack is incorporated within gNB as described above.

Now we briefly describe how a data path is established between UE and data network within a 3GPP 5G network. When a UE wants to connect with a data network, such as Internet, the 5G network establishes an end-to-end tunnel between UE and UPF (i.e., a Protocol Data Unit (PDU) Session). Additionally, a unique signaling link between UE and 5G Network (both at 5GC and RAN) is also established to exchange control messages between the UE and the 5G network. The end-to-end data tunnel takes the form of a Data Radio Bearer (DRB) at the air interface, while the signaling link takes the form of a Signaling Radio Bearer (SRB). Radio bearers are essentially layer 2 tunnels. Between gNB and 5GC (AMF+Session Management Function (SMF)), UE specific signaling messages

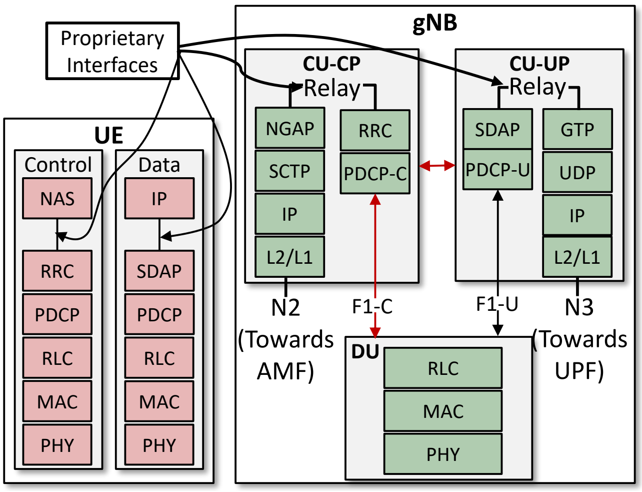


Figure 3.2: 3GPP 5G gNB and UE protocol stack

(NAS messages) are exchanged through a (per UE) unique association over N2 interface, whereas a PDU Session takes the form of a GTP tunnel between gNB and 5GC (UPF).

In 3GPP 5G network, different unique identifiers are used to identify UE specific signaling association and data over different interfaces. For example, to uniquely identify UE over N2 interface, gNB uses RAN UE NGAP ID, whereas AMF uses AMF UE NGAP ID. Similarly, UE specific data sessions are uniquely identified via GTP Tunnel End-point Indicator (TEID) on N3 interface.

3.3 Frugal 5G Implementation Challenges

We now discuss the challenges concerning the implementation of desired features proposed by Frugal 5G network architecture within 3GPP 5G framework. We discuss implementation challenges regarding three desired network characteristics, i.e., unified multi-RAT access network, unified interworking function, and flexible fog deployment. We do not discuss the *multi-operator resource sharing* as 3GPP 5G network architecture supports network slicing, which can be exploited to enable resource sharing among operators. To enable *localized communication*, a significant challenge is to provide user-level information in the fog element. We have not addressed this challenge in this thesis but aim to address it in future work.

1. *Unified multi-RAT access network*: As discussed before, 5GC exposes a common interface towards RAN irrespective of the RAT being used. The unification of RATs

happens at the core network level in 3GPP 5G network, while Frugal 5G network proposes a unified multi-RAT access network. To manage data traffic for a UE using multiple RATs concurrently, Access Traffic Steering, Switching & Splitting (ATSSS), an optional feature, has been introduced in 3GPP 5G specifications [32, 81]. An operator can define policies that direct UE and UPF (the data plane function in 5GC) to use multiple access. However, as 5GC handles the ATSSS feature, it may not be able to optimally perform traffic steering and load balancing across multiple RATs since it is unaware of RAN-level information such as traffic load and radio channel conditions.

2. *Unified interworking function*: The Frugal 5G network architecture proposes a unified interworking function. However, 3GPP 5G network defines RAT-specific interworking entities and associated interfaces. The underlying reason for RAT-specific interworking entities is the tight coupling between radio and core network protocol stack. As observed in Figure 3.2, the NGAP-RRC and SDAP-GTP protocol interfaces are tightly coupled with each other and utilize proprietary vendor specific communication mechanism. Though not shown in the figure, N3IWF also uses proprietary interfaces for communication between their radio and core network stacks. Similarly, higher layers (NAS and IP Layers) at UE are tightly coupled to the underlying radio protocol stack, leading to close interworking between RAN and 5GC. 5GC also becomes complex as registration state per RAT is maintained at 5GC if a UE is connected to more than one RAT.

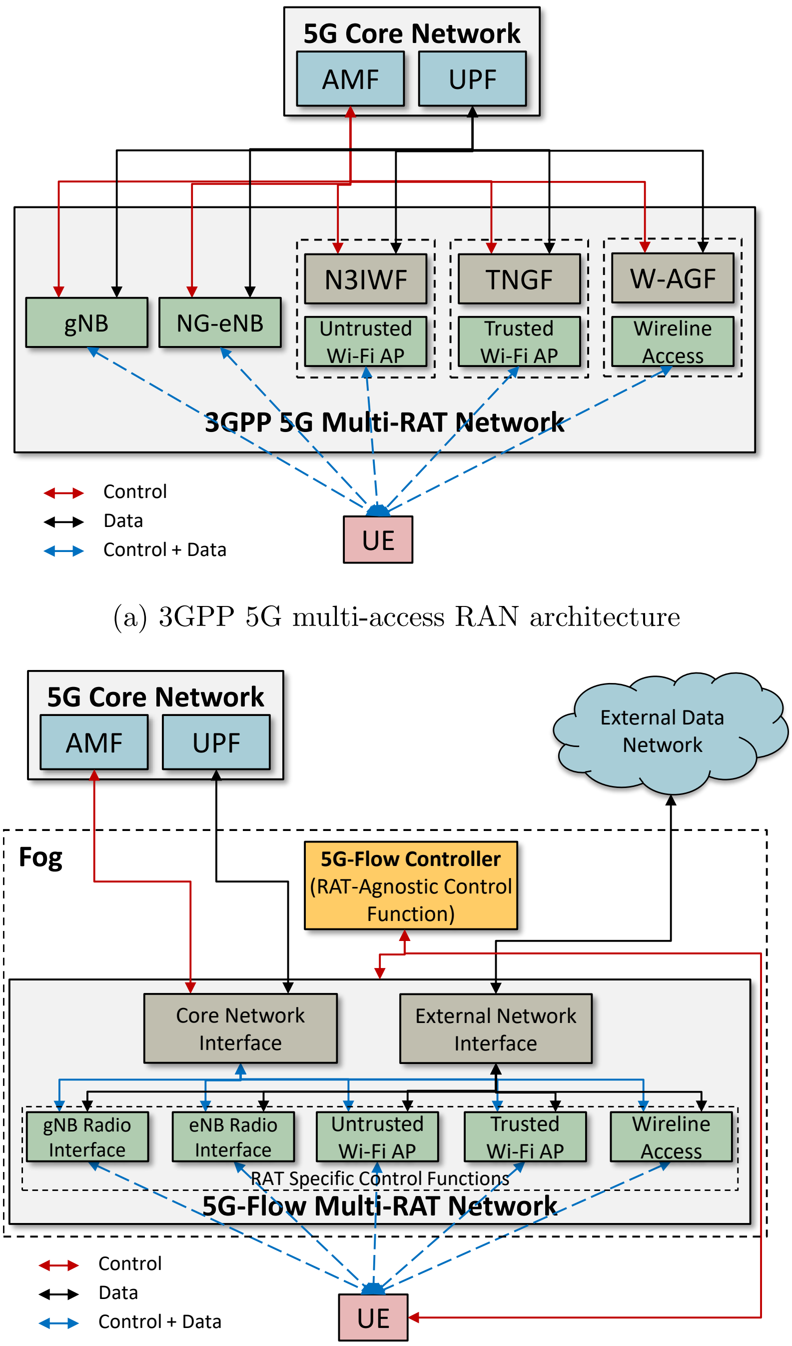
   Another implication of the tight coupling between RAN and 5GC is the lack of flexibility in 3GPP 5G network architecture. For instance, it is not possible to use 5G NR based gNB to connect with the Internet directly without routing the data through 5GC. Similarly, gNB can not be directly used with 4G core network in the 3GPP 5G network.

3. *Flexible fog deployment*: The current 3GPP 5G network does not support storage and compute capabilities at the access network. Therefore, to enable flexible fog deployment, it is required to introduce a fog element in the access network, enabling storage and compute capabilities.

3.4 5G-Flow RAN Architecture

The current 3GPP 5G multi-RAT RAN architecture is illustrated in Figure 3.3a. The figure shows how various access technologies interface with 5GC using separate interworking entities such as N3IWF, TNGF and W-AGF. Figure 3.3b illustrates our proposed multi-RAT RAN architecture, referred to as *5G-Flow*, to realize the desired Frugal 5G network characteristics. The *fog* element comprises a multi-RAT 5G-Flow RAN that communicates with 5GC through a *unified interworking entity* instead of separate interworking functions. Additionally, the interface between RAN and core network is flexible, such that any RAN can connect with any cellular core network or Internet directly. The software-defined *5G-Flow controller* acts as a *RAT agnostic control function* (defined under Frugal 5G network architecture in Section 2.3) and manages multiple RATs as shown in Figure 3.3b. The 5G-Flow controller is light-weight as it essentially supplements the core network functionality and performs specific and limited functions at the RAN. We use the existing radio protocol stacks of multiple RATs (without any modifications) as *RAT specific control functions*. The controller also manages the unified interworking entity that communicates with 5GC as well as dataflows across multiple RATs in RAN. Since the controller has access to RAN-level information such as traffic load and radio channel conditions, it can efficiently manage the downlink dataflows across RATs. As shown in Figure 3.3b, the controller also controls the UE, which enables uplink dataflow management in a multi-RAT RAN.

To realize the proposed 5G-Flow RAN architecture, we apply the OpenFlow concepts [77]. We envision 5G RAN as an OpenFlow network comprising a 5G-Flow controller (as OpenFlow controller) and OpenFlow switches that are instantiated on the network side and the associated UEs, as shown in Figure 3.4a. We have also introduced other required network functions under *RAN storage and compute* such as DHCP server, Authentication Server and Content Server in the fog. These network functions can be added or removed based on the requirements of rural areas to enable a *flexible fog deployment*. We now discuss the enhancements to the current 3GPP 5G architecture to realize the 5G-Flow network.



(a) 3GPP 5G multi-access RAN architecture

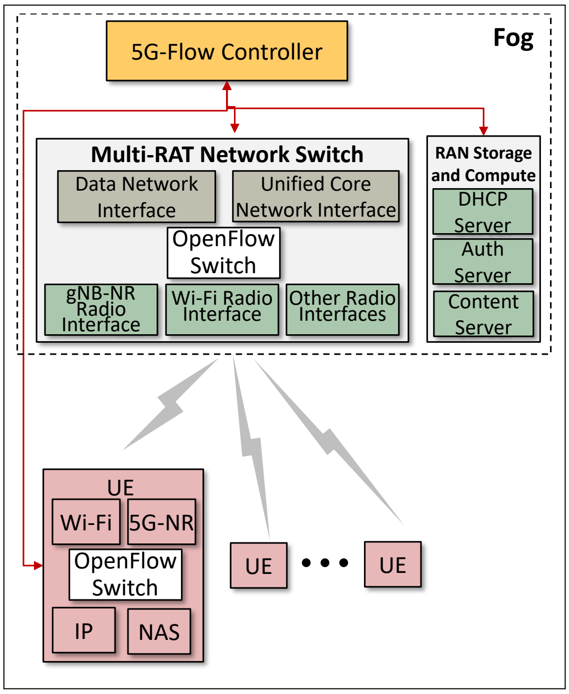
(b) Proposed 5G-Flow RAN architecture

Figure 3.3: Conceptual diagram for 5G-Flow RAN architecture and its comparison with the current 3GPP 5G RAN (radio interfaces, shown in green, represent both the control and data plane entities of the network)

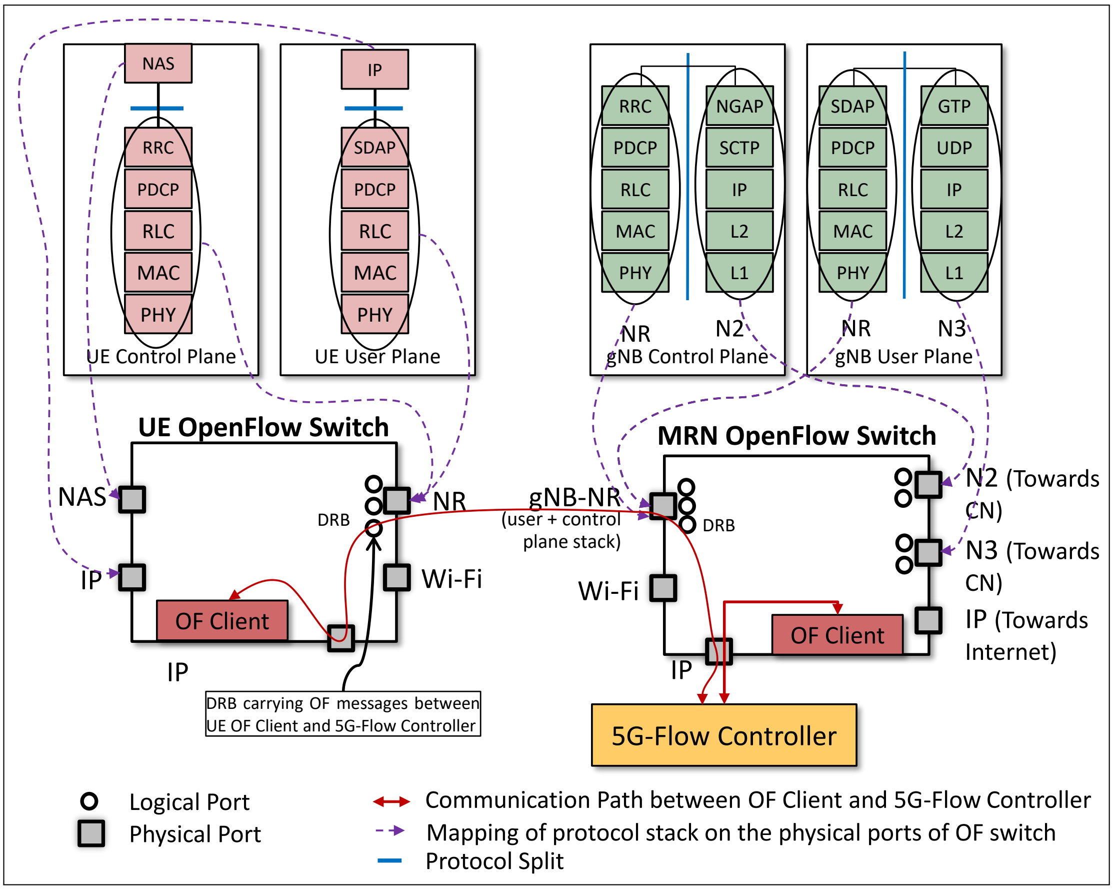
3.4.1 Proposed Multi-RAT Network

The existing 3GPP 5G RAN consists of various multi-RAT network nodes, including 3GPP access (e.g., gNB, eNB) and non-3GPP access (e.g., Wi-Fi, N3IWF). To integrate multiple RATs in 5G-Flow RAN and enable a unified interworking entity, we propose a "protocol split" between radio interface protocol stack and N2/N3 protocol stack of RAN nodes. For 3GPP access nodes such as gNB, split happens at the gNB node itself, whereas for non-3GPP access, it is done at the interworking function such as N3IWF. To illustrate the protocol split, we take an example of gNB, as shown in Figure 3.4b. A gNB consists of NR protocol stack (which interfaces with a UE) and N2/N3 protocol stack (which interfaces with the 5GC). As shown in the figure, we split gNB vertically and separate gNB-NR and N2/N3 protocol stack. We can similarly split the N3IWF node. We now introduce an OpenFlow switch, referred to as Multi-RAT Network (MRN) OpenFlow switch, which is responsible for bridging radio and N2/N3 protocol stacks of multiple RATs. These protocol stacks form different interfaces (physical ports) of the MRN OpenFlow Switch. The 5G-Flow controller directs MRN OpenFlow switch to process the messages from different RATs and deliver it to N2/N3 protocol stack towards 5GC and vice-versa. This way, the MRN OpenFlow switch along with the 5G-Flow controller replaces all RAT-specific entities such as gNB, eNB, N3IWF, etc. and exposes a unified interface towards core network.

The MRN OpenFlow switch has physical ports both at the radio and 5GC interface side, as shown in Figure 3.4b. Both the control plane (RRC and underlying protocol stack) and the data plane (SDAP and underlying protocol stack) of gNB-NR radio interface map to one of the radio side ports of the OpenFlow switch. Similarly, Wi-Fi MAC and physical layer map to another port. NGAP and GTP protocol layer (along with underlying N2/N3 stack) map to the physical ports on the 5GC side. The physical port, labeled as IP, enables a UE to directly connect with Internet without engaging 5GC. As the radio interface protocol stack is independent of N2/N3 protocol stack, OpenFlow switch can steer the data traffic of a UE towards IP port, enabling direct connectivity with Internet. An interface towards 4G core network is not shown in the figure, but it can be easily incorporated by adding S1-Mobility Management Entity (MME) interface as a separate



(a) Block Diagram of 5G-Flow RAN



(b) Application of OpenFlow Protocol at RAN

Figure 3.4: Implementation of 5G-Flow RAN architecture using OpenFlow protocol

physical port in the proposed OpenFlow switch. This feature enables a UE with 4G compatible NAS layer to communicate with 4G core network via 5G RAN.

As shown in Figure 3.4b, Wi-Fi MAC layer is mapped to the radio side port in MRN OpenFlow switch. However, the user plane of N3IWF uses additional protocol layers i.e., Generic Routing Encapsulation (GRE) and IP Security (IPsec), for creating a secure tunnel between UE and N3IWF over Wi-Fi radio interface. Similarly, it uses TCP and IPsec protocol layers in the control plane for encapsulating NAS messages. Our architecture provides flexibility in employing these protocols. If a UE does not want to use GRE and the underlying protocol stack for some dataflows, a logical port can be created at UE and RAN, which transparently passes the data packets through Wi-Fi interface without any processing. Further, a logical port enabled with GRE, and IPsec protocol layers can be created over Wi-Fi MAC based physical port if a UE needs a secure tunnel for another data flow.

3.4.2 Proposed Enhancements at UE

An OpenFlow switch, introduced at UE, decouples NAS (that communicates with 5GC) and IP protocol layers from underlying radio protocol stack, as shown in Figure 3.4b. We introduce a common IP layer instead of RAT-specific IP layers. There can be different radio interface protocol stacks depending on the technology, but NAS and IP layers remain common. At the radio side ports of UE OpenFlow switch, NR (RRC/SDAP, and underlying protocol stack) and Wi-Fi (MAC and physical layer) radio stacks are mapped.

The UE OpenFlow switch, along with the 5G-Flow controller, manages the UE's radio connectivity and enables uplink dataflow management across multiple RATs. Moreover, when a UE is connected to 5GC via multiple RATs, it registers only once. Currently, a UE connected to 5GC through more than one RAT has to register with 5GC separately via each RAT. With the separation of NAS layer from radio protocol stack (in addition to radio and N2/N3 protocol split at the network side), UE's communication with RAN is completely decoupled from its communication with 5GC. This feature allows a UE to flexibly connect to different networks such as 5GC, 4G core network, or directly to Internet.

UE1 OpenFlow Switch

	Rule	Action
NAS	Incoming NAS message at **LP1** (SRB)	Forward to **NAS**
NAS	Incoming packet at **NAS**	Forward to **LP1** (SRB)
Data	Incoming data packet at **LP2** (DRB)	Forward to **IP**
Data	Incoming packet at **IP**	Forward to **LP2** (DRB)

MRN OpenFlow Switch

Rule	Action
Incoming NAS (UL) message at **LP1** (through RRC on SRB)	Forward to **LP5**
Incoming NAS (DL) message at **LP5** (through NGAP)	Forward to **LP1**
Incoming data packet (UL) at **LP2** (DRB UE1)	Forward to **LP7**
Incoming data packet (DL) at **LP7** (GTP UE1)	Forward to **LP2**

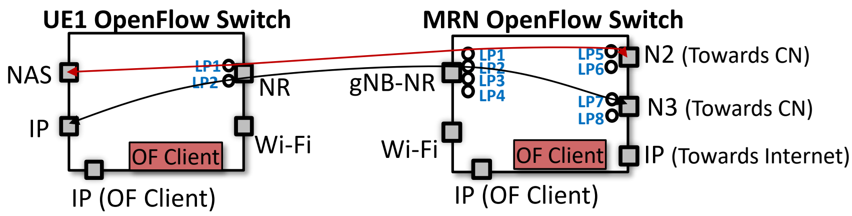


Figure 3.5: Example of flow entries at UE and MRN OpenFlow switch for NAS signaling and data transfer. Figure shows how Logical Ports (LPs) are used to set up a flow in the proposed network.

3.4.3 OpenFlow Switch Configuration

As discussed before, OpenFlow controller can create and configure logical ports on an OpenFlow switch using OF-Config protocol. The meaning (processing) associated with these logical ports varies according to the underlying interface being used. On the gNB-NR interface at MRN OpenFlow switch, a logical port represents a UE-specific radio bearer. At least two logical ports are created for each UE, one for SRB and another for DRB. To identify these logical ports uniquely, we use SRB/DRB ID as defined in 5G specifications. A logical port at NGAP interface signifies a UE-specific NGAP association and is identified by RAN UE NGAP ID. Similarly, at GTP interface, a logical port implies a PDU session of a UE and is identified by a GTP-TEID.

At NR interface of the UE OpenFlow switch, a logical port represents SRB and DRB of a UE (similar to MRN OpenFlow switch). The logical port representing SRB processes the RRC and underlying protocol stack, while the one representing DRB processes SDAP and underlying protocol stack. At IP interface, logical ports signify ongoing PDU sessions for a UE.

Despite the differences in processing at each of the physical interfaces, usage of logical ports provides a uniform abstraction to be used by 5G-Flow controller to configure flow paths through the switch. The 5G-Flow controller simply configures logical ports on physical interfaces in an OpenFlow switch. It is the responsibility of the interface to translate OF-Config message (for port creation) to interface specific handling. For instance, a DRB, along with a GTP tunnel, needs to be established during a PDU Session setup. When the gNB-NR interface receives a message to create a new logical port corresponding to a DRB, the RRC layer on the gNB-NR interface translates it to configure its underlying lower layers, e.g., PDCP, RLC layers for the local DRB configuration. It also exchanges *RRC Reconfiguration* messages with the UE for a corresponding DRB configuration on the UE. Similarly, a logical port creation message sent to N3 interface is translated (by the interface) to create a GTP tunnel. A logical port creation message to N2 interface gets translated into the creation of a unique UE-specific NGAP association with 5GC.

Once the logical ports are created on the interfaces, 5G-Flow controller defines a mapping between different logical ports across the interfaces on the OpenFlow switch, as shown in Figure 3.5. For instance, a mapping between DRB and GTP tunnel is created for a UE, or a mapping between UE-specific NGAP association and SRB is created. The controller realizes these mappings through flow entries added at the OpenFlow switches. These port mappings enable the simple forwarding of data and signaling messages through the OpenFlow switch and also makes the control and management task for the 5G-Flow controller easy.

A key novelty of the proposal is the usage of GTP tunnels, radio bearers, or UE-specific NGAP association as logical ports. These entities (a radio bearer, or a UE-specific NGAP association) carry a specific set of data flows or messages. By using them as logical ports, we can virtualize them and enable their manipulation by an OpenFlow Controller through OpenFlow protocol.

3.4.4 Impact on 3GPP 5G Architecture

It is important to note that the 5G-Flow architecture proposes solutions in areas that 3GPP 5G specifications have left open for vendor implementation. For example, 3GPP has not defined specific interface between the RRC and NGAP stack on the gNB-CU or the interface between the NAS and the RRC layers on the UE. Our work achieves the unification of multiple RATs in the access network by standardizing such interfaces with the help of OpenFlow. Moreover, the proposed architecture is substantially software-based and can be easily implemented. Our proposal requires instantiation of OpenFlow switches at UE and MRN. Even though gNB-CU and N3IWF has been replaced with MRN Switch and 5G-Flow Controller, the changes are more from the organization perspective. We use the existing networking concepts given by OpenFlow protocol such as logical ports to enable a uniform abstraction to be used by the 5G-Flow controller. The processing of RRC and underlying protocol stack or NGAP and underlying protocol stack is done as part of logical port processing. The 5G-Flow controller need not know the working of these protocol stacks, thereby making its implementation relatively simple.

The proposed software-related changes do not modify the existing protocols/interfaces of the 3GPP 5G architecture. The interfaces such as N2/N3 interface between RAN and 5GC, radio interface between UE and RAN, and F1 interface between gNB-CU and gNB-DU may continue to be used without any modifications under the 5G-Flow framework.

3.5 Working of the 5G-Flow RAN

In this section, we describe some important system procedures to explain the working of 5G-Flow network. We also discuss how 5G-Flow network enables direct connectivity with the Internet without involving core network. Additionally, we discuss how 5G-Flow improves dataflow management in the access network.

3.5.1 Initial Connection Setup

The functioning of 5G-Flow network requires communication between OpenFlow switches and 5G-Flow controller. The control messages between OpenFlow switch and the controller are exchanged over a TCP/IP connection. The MRN OpenFlow switch and 5G-Flow controller are co-located and can communicate over a wired interface. Figure 3.6 shows how communication between UE OpenFlow Client (UE OpenFlow switch) and 5G-Flow controller is established. Here, we have assumed that the default path for UE-controller communication is via gNB-NR interface, but the path can be established via

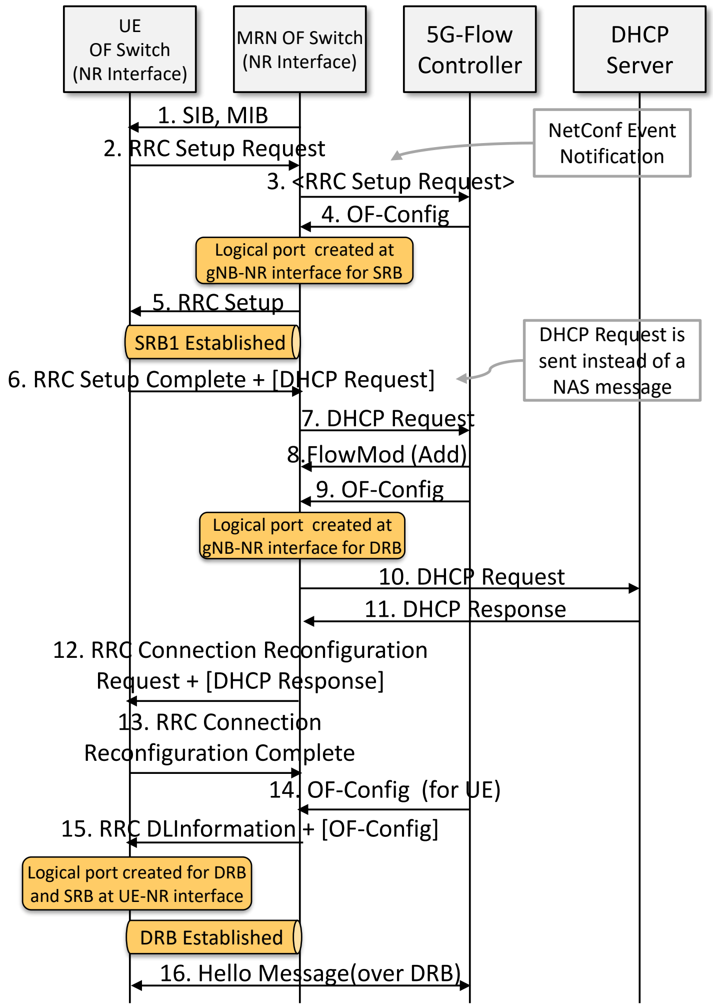


Figure 3.6: Initial connection setup

Wi-Fi radio interface as well. We explain the call flow to set up the initial connection next.

- UE receives system information and sends an *RRC Setup Request* to establish a radio connection with gNB-NR interface. Since 5G-Flow controller is responsible for admission control, it notifies the controller via a NetConf notification so that a decision for UE admission can be taken. If the UE is granted access to the network, the 5G-Flow controller sends an OF-Config message to create a logical port on gNB-NR interface for the subsequent signaling messages exchange (via SRB1) with the UE. The gNB-NR interface sets up SRB1, maps it to the logical port, and sends *RRC Setup*. As the UE OpenFlow client does not have a TCP/IP connection with the controller, a logical port can not be created at UE using OF-Config. Instead, UE uses the default physical port mapped to 5G-NR interface for initial signaling.

- UE responds with *RRC Setup Complete* message and a *DHCP request* is sent in NAS message field instead of *Registration Request*. gNB-NR forwards the encapsulated NAS message towards the 5G-Flow controller as the flow entry for a NAS message from the UE is not present in MRN OpenFlow switch. When the controller receives this message, it sends a FlowMod (Add) command to add the flow entry at MRN OpenFlow switch. 5G-Flow controller also sets up a logical port at gNB-NR interface for DRB of the UE, using OF-Config message. This DRB is created for carrying OpenFlow client messages from UE to the 5G-Flow controller.

- After the flow entry addition, gNB-NR interface forwards DHCP request to the DHCP server. When DHCP response is received, it is sent in the NAS message field of *RRC Connection Reconfiguration Request*. This message also informs UE to set up a DRB and confirm through *RRC Connection Reconfiguration Complete* message.

- An OF-Config message is sent to UE to configure a logical port and map it to the newly established DRB. This message creates a logical port for SRB as well, which is used for future signaling messages. The communication path between UE OpenFlow switch and the 5G-Flow controller is now established via DRB, and they exchange *Hello* message over this DRB.

This call flow illustrates how UE's communication with RAN has been decoupled with its communication with core network. UE can use RAN to exchange DHCP and OpenFlow messages with entities located in the edge instead of exchanging NAS messages and data with the core network.

3.5.2 Registration and PDU Session Setup with the 5G Core Network

If a UE wants to access the 5G cellular network, it needs to register with the 5GC. Here, we take an example of how a UE registers via gNB. The registration of a UE can also take place via a Wi-Fi network, if available. To facilitate UE's communication with the

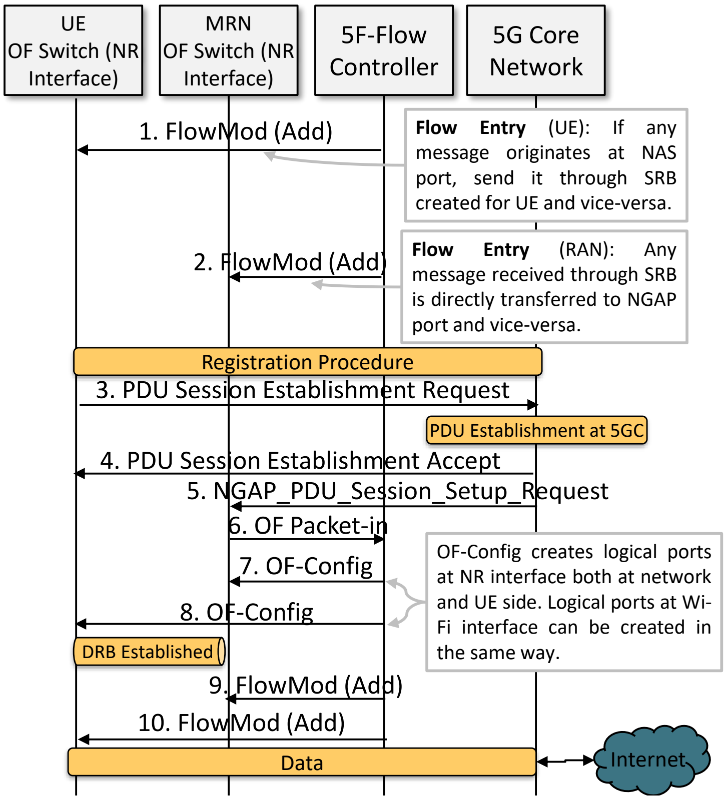


Figure 3.7: Call Flow to illustrate the communication of UE with 5G Core Network

5GC, 5G-Flow controller can proactively set up a path that delivers NAS messages from UE to the 5GC and vice-versa, as shown in steps 1 and 2 in Figure 3.7. It can also be implemented through a reactive method in which the 5G-Flow controller adds the flow entries after the first NAS message originates at UE, and a table-miss is detected. Next, we discuss how a PDU session is established via 5G-Flow RAN.

- When AMF (in 5GC) receives the NAS message (*PDU Session Establishment Request*) from a UE, it informs SMF, which creates a PDU session for the UE. SMF sends PDU session related information to AMF, i.e., GTP tunnel end-point and QoS information. AMF forwards this information as an NGAP message to (R)AN along with a NAS message (*PDU session Establishment Accept*).

- The MRN OF switch transparently forwards the NAS message to the UE. However, unlike NAS message, it processes the NGAP message and forwards PDU session related attributes to the 5G-Flow controller via *OF Packet-in* message. For processing the NGAP message and enabling communication between gNB-NR interface stack and 5G-Flow controller, a new protocol definition may be required. However, the protocol will be required for limited tasks such as processing and exchanging PDU Session (data bearer) related parameters received/sent via NGAP messages, while most tasks are handled via logical port processing in OpenFlow switch.

- The controller, based on the Packet-in message and RAT-specific information, decides how the incoming dataflow should be distributed among the available RATs. It, then, sends OF-Config messages to the radio interfaces to create logical ports, which are interpreted by these interfaces for configuration of the underlying protocol stack. For instance, configuration of lower layers by RRC to create a DRB when a logical port creation message is received at gNB-NR interface. Also, an OF-config message is sent to GTP interface of the MRN OF switch for creating a logical port which signifies GTP tunnel for a UE's PDU Session.
- After this, flow entries are added at the UE OF switch and the MRN OF switch that maps the newly created DRB to IP port at UE and maps the DRB to the GTP tunnel, respectively. The path for UE PDU session is now set up.

3.5.3 Direct Connectivity to the Internet

A UE in a cellular network can access Internet solely through the core network. It does not have the flexibility to connect to the Internet from RAN bypassing the core. This feature can be beneficial in rural areas where users are mostly stationary, and a complex core network may not be required. However, the advantages of the cellular stack with an efficient L2/L1 layer can still be exploited. Figure 3.8 illustrates the call flow of how a UE in the 5G-Flow network can access the Internet directly without going through 5GC. Unified interworking function is the key characteristic enabling this feature.

- UE sends a packet to the IP interface of the UE OpenFlow switch intending to connect directly with the Internet. When the switch observes a table miss as the flow entry for this packet does not exist, it forwards the packet to the 5G-Flow controller via *Packet-in* message.
- Depending on the QoS requirement, the flow controller decides whether a dedicated DRB needs to be created. In case a dedicated DRB is required, the flow controller

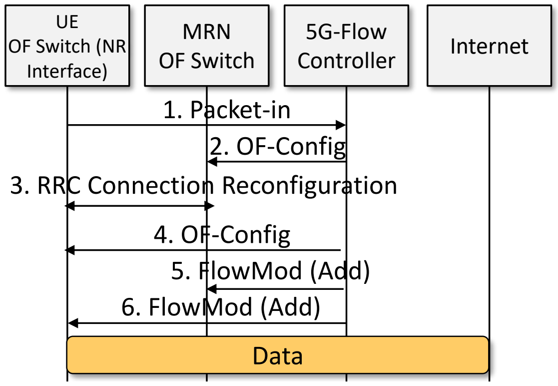


Figure 3.8: Call Flow to illustrate the communication of UE with the Internet

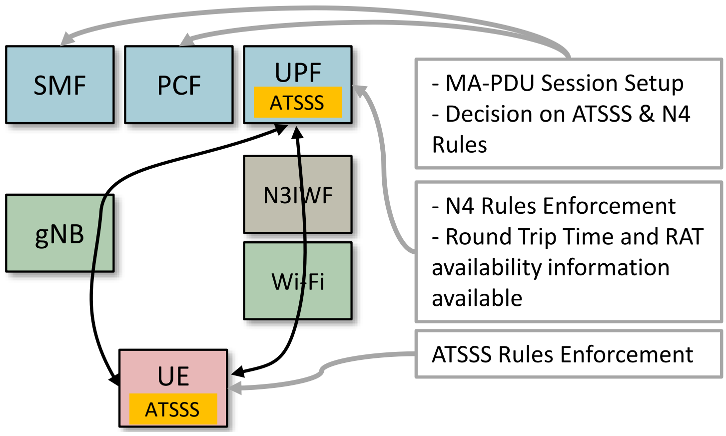
sends an *OF-Config* message to gNB-NR interface to create a logical port, which is translated by RRC layer and a DRB is created using *RRC Reconfiguration* messages.

- An OF-Config message is also sent to UE so that it can create a logical port on the NR interface and map the newly established DRB to the logical port. The 5G-Flow controller adds the appropriate flow entries at UE and MRN OpenFlow switch so that UE can access the Internet.

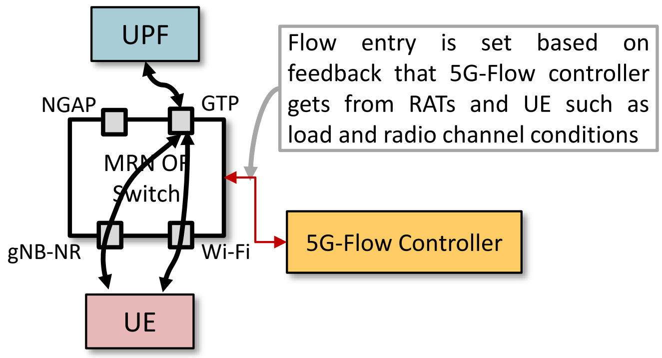
3.5.4 Dynamic Dataflow Management

5G has introduced Access Traffic Steering, Switching & Splitting (ATSSS) feature, which manages the multi-access traffic in the 5G network. Figure 3.9a illustrates how the multi-access traffic flows through the 5G network [32, 81]. UE can initiate a Multi-Access-Protocol Data Unit (MA-PDU) session to enable PDU exchange between UE and UPF via 3GPP and non-3GPP RATs simultaneously. To manage the uplink traffic, UE considers the ATSSS rules provided by the 5GC. To manage the downlink traffic, UPF considers ATSSS rules provided by SMF over N4 interface along with the feedback information from UE. However, the feedback information available at UPF is limited, i.e., Round Trip Time (RTT) and RAT availability. This information may not be sufficient to route the flow through multiple RATs optimally.

As shown in Figure 3.9b, the flow configuration in the 5G-Flow network happens at the RAN-level, where utilizing various RAT-specific attributes is viable. The controller can periodically access the value of specific attributes, such as traffic load at RATs, flow



(a) Dataflow in 5G Network under ATSSS feature



(b) Dataflow in 5G-Flow Network

Figure 3.9: Comparison of dataflow with 5G ATSSS feature and dataflow with 5G-Flow architecture

statistics, and radio signal strength, to optimize the dataflow. The first two attributes are available at the OpenFlow switch. To access the radio signal measurement available at RRC (gNB-NR), the controller can subscribe to the measurement data at the OpenFlow switch via NETCONF notifications [80]. The asynchronous notification mechanism is supported by OF-Config protocol and it allows us to define notification messages, in addition to the already available set of notifications.

Based on these parameters, an optimized policy for flow management can be determined, and the incoming flow is split across available RATs, as shown in the figure. We also have the flexibility to decouple uplink and downlink in 5G-Flow RAN. As the controller manages MRN and UE OpenFlow switch separately and can add distinct flow entries in both the switches, uplink and downlink for a user can be easily decoupled. We

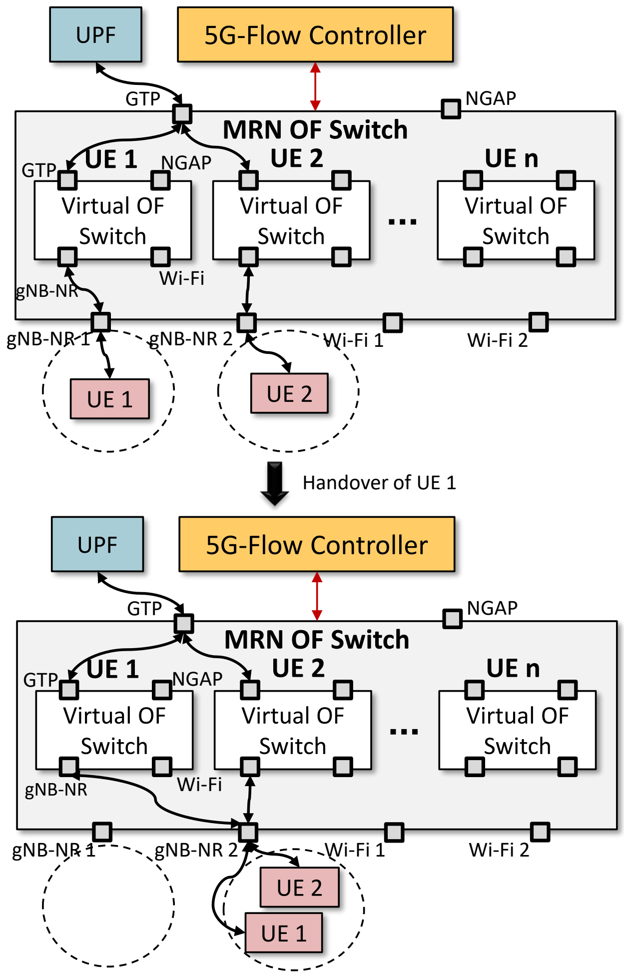


Figure 3.10: Mobility Management

evaluate the performance of data flow management in the 5G-Flow network viz-a-viz the 3GPP 5G network in Chapter 4.

3.5.5 Mobility Management

To decrease the load at 5GC, it is essential to reduce the handover signaling between 5GC and RAN. In the 5G-Flow network, the controller can handle multiple gNB-NR and Wi-Fi interfaces as shown in Figure 3.10 (Section 3.4). The handover between two gNB radio interfaces under the same controller is handled by the controller without involving 5GC. To simplify the handover procedure at the RAN, we propose a novel idea wherein an OpenFlow switch per UE is instantiated (using OF-Config) over the MRN OpenFlow switch. An OpenFlow switch per UE is not required for stationary users. It holds the ongoing data session context and RAT connectivity information for a UE. As shown in the figure, there is a mapping from the physical ports to the logical ports of the virtual OpenFlow switches for each UE.

In 5G-Flow network, a mobile UE periodically sends radio signal measurement reports to gNB-NR. To facilitate handover (if required), we define an event based on the threshold signal strength required for handover. The 5G-Flow controller, which has subscribed to this event notification, takes a handover decision based on the radio signal strength and other network parameters such as load on RATs. If a handover decision is taken, the data path for a UE at the radio side needs to be changed, while there is no change in the data path at the core network side. As shown in Figure 3.10, gNB-NR port of virtual OpenFlow switch of UE 1 now maps to gNB-NR 2 (instead of gNB-NR 1) after the handover. The virtual OpenFlow switch remains unaffected, and hence managing UE mobility becomes very simple. All the ongoing sessions of the UE remain intact. Moreover, this handover is completely transparent to the 5GC as connectivity towards UPF remains unchanged i.e., no modifications are required in GTP tunnel.

3.6 Additional Use-cases

Apart from serving as a solution for rural broadband, there are some additional use-cases supported by 5G-Flow network architecture which are discussed next.

3.6.1 Captive Network

Captive networks such as public Wi-Fi networks are not limited to business or campus environments but are increasingly deployed in stadiums, malls, bus-stops, and metro stations. These captive networks suffer from coverage holes. To solve the problem of coverage holes and enable seamless connectivity, we can integrate macro gNB and Wi-Fi AP in a captive network. A smaller number of Wi-Fi APs will be required to provide similar coverage through gNB and Wi-Fi networks.

Existing cellular radio technologies, e.g., 4G LTE or 5G NR, can not work in a standalone manner without core network. It is possible to include cellular radio technology in a captive network only if the core network is also present there. Usage of core network increases the cost of deployment for captive networks. One of the key advantages of the 5G-Flow network architecture is that it allows usage of cellular radio technology (5G

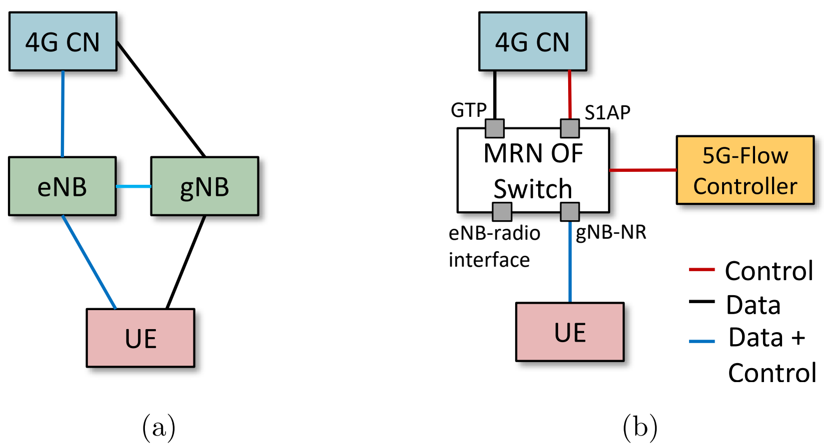


Figure 3.11: Implementation of non-standalone 5G architecture proposed by (a) 3GPP, viz-a-viz (b) 5G-Flow Network

NR/4G LTE) without involving the core network. In a 5G-Flow based captive network, a UE can use 5G NR to directly connect to a private data center of the captive network or Internet without going via core network. Core network is not required for initial access or authentication. A simpler authentication mechanism, as deemed appropriate, can be built over the 5G-Flow network and deployed in edge as part of *RAN compute and storage*.

3.6.2 Simpler Mechanism for 5G Non-standalone Deployment

It appears that 5G deployments will be carried out in phases, and the early adopters of 5G are most likely to choose a non-standalone deployment. This method involves deployment of eNB and gNB in RAN, which communicate with the 4G core network as shown in Figure 3.11. A UE should support dual connectivity to both eNB and gNB to avail 5G services. Since gNB does not communicate with 4G core network, it is modified to communicate with 4G core network and referred to as en-gNB. An eNB acts as a master node while en-gNB acts as a secondary node, and they communicate via X2 interface. All signaling exchange between UE and the network happens via eNB, while gNB is only used for data transfer. The non-standalone method is a much faster way to deploy 5G as it leverages the existing 4G infrastructure. However, 5G RAN capability can not be exploited entirely in the current architecture.

5G-Flow RAN architecture enables non-standalone deployment in a much simpler manner. Due to the complete decoupling between RAN and core network connectivity, it allows a UE to use 5G NR interface to connect to 4G core network and receive mobile data services without necessarily using the dual connectivity. At the same time, 5G-Flow also enables a UE to use dual connectivity, whenever required.

3.7 Summary

In this chapter, we presented 5G-Flow RAN architecture that intends to realize the desired characteristics of Frugal 5G network within 3GPP 5G framework. Towards this, the 5G-Flow RAN supplements the 3GPP 5G network with OpenFlow. Following are the key features of the 5G-Flow network.

- We propose an OpenFlow based unified and integrated multi-RAT RAN architecture. It has eliminated the use of multiple interworking functions and associated interfaces used in the current 3GPP 5G architecture.

- The OpenFlow switches instantiated in the RAN along with the 5G-Flow controller help manage the dataflows efficiently both in downlink and in uplink direction. The 5G-Flow controller is aware of RAT-specific attributes and hence can take efficient flow management decisions. In addition to enhancing the user experience, it also helps in efficient radio resource utilization.

- A light-weight 5G-Flow controller enables RAN level management of downlink as well as uplink dataflows, in turn, utilizing the multi-RAT resources efficiently.

- Decoupled radio and core network protocol stacks at RAN nodes and separation of NAS layer from the radio protocol at UE fully decouples UE's connectivity to RAN from its connectivity to core network. This brings immense flexibility and enables a UE to interface with 5GC, 4G core network, Internet, or any other data network via any 4G/5G/Wi-Fi based RAN.

- The proposal requires minimal changes in RAN. Moreover, these changes are software-based and can be easily implemented. No changes are proposed in the protocol used between different network entities, i.e., UE-gNB, UE-5GC, or gNB-5GC interface.

- With the introduction of a virtual OpenFlow switch for every UE, the mobility management at RAN can be easily handled. Since there is no modification in GTP connection towards UPF during handover, the signaling load towards 5GC decreases.

The above-stated features of 5G-Flow RAN meet some of the essential characteristics of Frugal 5G network, i.e., unified multi-RAT network, unified interworking function, flexible fog deployment and multi-operator resource sharing. However, to enable localized communication, further enhancements are required in the 5G-Flow network. We briefly discuss these enhancements in Chapter 7 as part of the future work of this thesis. In the next chapter, we discuss the performance analysis of the 5G-Flow network.

Chapter 4

Quantitative Assessment of the 5G-Flow Network

The detailed architecture and working of the 5G-Flow network have been discussed in the last chapter, suggesting a practical realization of the Frugal 5G network. Specifically, the proposed 5G-Flow network has enabled a framework that can efficiently manage a multi-RAT RAN via 5G-Flow controller. In this chapter, we focus on analyzing the performance of data flow management across multiple RATs in a 5G-Flow network. We evaluate the performance of the 5G-Flow network using system-level simulations. An open-source 5G simulator with a fully developed protocol stack is not yet available, so we develop a 5G simulator supporting for multiple RATs. In our simulator, we have implemented physical and MAC layer protocol stacks for 5G-NR and Wi-Fi RATs. We have also implemented 5G-Flow controller that manages these RATs.

We evaluate the performance of downlink and uplink dataflow management in the 5G-Flow network and compare it with existing 3GPP 5G network performance. To examine the downlink dataflow management, we consider a network that offers different data services to UEs which differ in their priorities. We propose a heuristic RAT selection algorithm that considers UE's service type, available RATs, and load at the RATs. The performance analysis of network throughput and delay shows that 5G-Flow outperforms 3GPP 5G network and improves load balancing across RATs. As every UE features an OpenFlow switch, the 5G-Flow controller can efficiently manage uplink dataflow and also decouple uplink dataflow from downlink at UE. The performance analysis demonstrates that the above-stated characteristic improves the uplink throughput and delay performance of the network.

The rest of the chapter begins by introducing the structure of the multi-RAT 5G-Flow simulator. We explain the various steps involved in the simulation process and describe network stack implementation (Section 4.1). Next, we discuss the performance analysis of downlink and uplink dataflow management, wherein we describe the proposed heuristic algorithm and compare its results with the 3GPP 5G network (Section 4.2). We finally summarize the chapter (Section 4.3).

4.1 Simulator Framework

In this section, we discuss the structure of the 5G multi-RAT system-level simulator (developed in MATLAB). We have released the source code of the 5G multi-RAT simulator under MIT License for the benefit of other researchers, and it is available online [45]. The main objective of developing the simulator is to test the performance of 5G-Flow network architecture. However, the simulator offers sufficient flexibility, facilitating the implementation and testing of different multi-RAT architectures. Packets are the fundamental objects of our network simulator. The simulator supports multiple RATs, i.e., gNB-NR and Wi-Fi, and we describe the implementation of physical and MAC layers of these RATs. We also discuss the implementation of the application and transport layer, which both the RATs share. Moreover, a centralized controller has been implemented that manages these RATs. The simulator supports downlink as well as the uplink communication channel. Figure 4.1 shows the overview of the important steps of the simulation process, and these steps are explained as follows.

- *Initialization*: This step involves initializing all the necessary parameters for both gNB-NR and Wi-Fi RATs. Based on the initialized parameters, network nodes comprising gNB-NR, Wi-Fi APs, and UEs are generated. Geographic locations are assigned to these network nodes, which determines the network layout. The network nodes remain stationary for the entire simulation. We aim to enhance the simulator to support mobility in future.

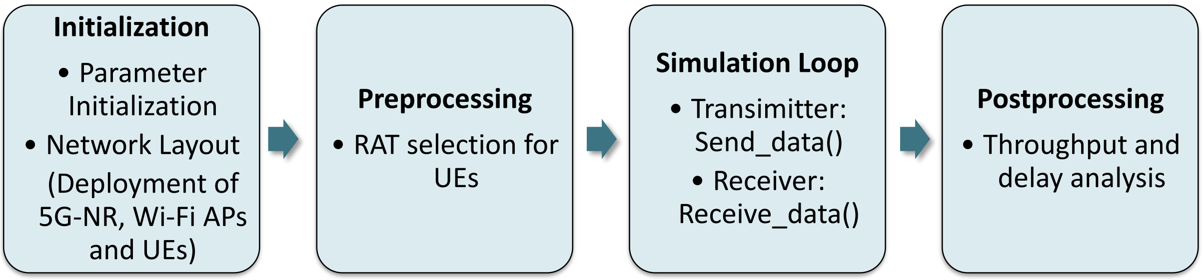


Figure 4.1: Simulator overview

- *Preprocessing*: Before simulation starts, a UE must be associated with a RAT(s). For this, we run a RAT selection algorithm. The RAT selected for a user remains constant for the entire simulation. We propose heuristic RAT selection algorithms for uplink and downlink dataflow management, and they are discussed in Section 4.2.

- *Simulation Loop*: We have developed a time slot based simulator, wherein the simulation timeline is divided into time slots. We use Time Division Duplex (TDD) that allows uplink and downlink transmission over the same frequency band but over different time slots. The uplink and downlink channels are quite similar, with minor differences in physical layer implementation. We have implemented a transmitter and a receiver process at the network side as well as at every UE. To transmit data at the network side on the downlink communication channel, the `send_data()` process is invoked. Similarly, `send_data()` process runs at every UE to send data on the uplink channel. The `send_data()` process implements the application, transport, MAC, and physical layers of both the RATs. The network stack implementation is explained in detail in the following subsection. In parallel, `receive_data()` function, at the network side and every UE, processes the data that has been successfully received.

- *Postprocessing*: Once the simulation is complete, all the key attributes such as TCP throughput of individual RATs, total TCP throughput, and packet delay are analyzed. We define packet delay as the total time it takes for a packet to reach its destination node from the source node.

4.1.1 Network Stack Implementation

We now describe the implementation of the network stack in the simulator at the network and every UE. A pictorial representation of network stack implementation in the multi-RAT simulator is given in Figure 4.2. We have not implemented IP layer as we have considered a simple network comprising a single macro cell.

Application Layer

The simulator supports VoIP and video streaming applications, while other applications can be easily added. We assume that the application layer is located in the access network, and the applications require a TCP connection to send the data to the destination node. We consider that the VoIP session generates packets after every 20 ms. The VoIP payload size is 160 bytes, and 45 bytes of overhead is assumed. We assume that the video streaming application generates packets following a homogeneous Poisson process with rate λ. Each packet is of size 1000 bytes. When an application generates a packet, it is added to the transmission buffer, subsequently accessed by the transport layer.

Transport Layer

At the transport layer, the simulator implements the TCP protocol. The `TCP_socket` extracts the packets from the transmission buffer located at the application layer. The `TCP_socket` maintains the congestion window and updates it depending on the ACKs received. Based on the congestion window, the packets are added to the TCP buffer. `TCP_socket` adds TCP header to the packets and sends it to the MAC layer, which is then transmitted by the physical layer. When a UE receives a packet on the downlink channel, the packet is inserted into the reception buffer. The packets are assembled and sent to `TCP_socket`, which generates ACK upon the packet's successful reception. The `TCP_socket` removes the header and sends it to the application layer.

MAC Layer

5G NR: The responsibility of the MAC layer is to schedule the TCP packets received from the transport layer. Under 5G NR MAC layer implementation, the scheduler allocates the Resource Blocks (RBs) to each user, wherein an RB is defined as 12 consecutive subcarriers in the frequency domain and 14 symbols in time domain [82]. The frame length for 5G-

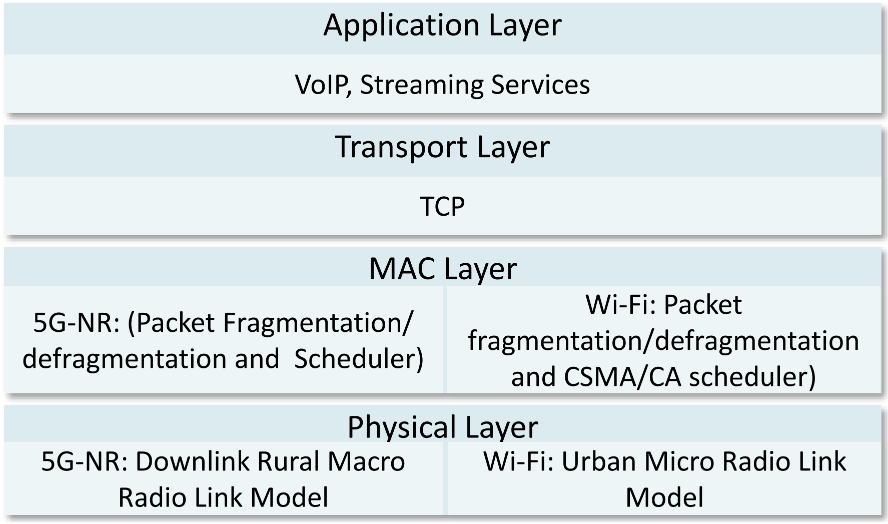


Figure 4.2: Network stack implementation in 5G Multi-RAT simulator

NR is 10 ms and the sub-frame length is 1 ms. 5G NR enables flexible numerology, i.e., flexible sub-carrier spacing and symbol length. The sub-carrier spacing configuration can be evaluated using $\Delta f = 2^{\mu} \times 15$ kHz, wherein μ can take values between 0 to 4. Let us understand 5G NR numerology with an example. For numerology $\mu = 1$ (30 kHz sub-carrier spacing), one resource block spans 360 kHz in frequency domain and 0.5 ms in time domain. The maximum number of resource blocks (N_{RB}) for a given channel bandwidth and numerology can be calculated as follows:

$$N_{RB} = \frac{(\text{Channel Bandwidth} - 2 \times \text{Guard Bandwidth})}{\text{One RB Bandwidth}}, \tag{4.1}$$

$$N_{RB} = \frac{(60 \times 10^3 - 2 \times 845)}{360} = 162 \text{ RBs}, \tag{4.2}$$

where,

Channel Bandwidth = 60 MHz,

Guard Bandwidth = 845 kHz ($\mu = 1$),

RB Bandwidth = 360 kHz ($\mu = 1$).

The simulator supports two types of schedulers, i.e., round-robin and priority-based scheduler [83]. In 5G NR, a slot is used as a dynamic scheduling unit and the RBs are scheduled in every slot. After scheduling, packets are fragmented based on the available Transport Block Size (TBS) of the allocated RBs. A transport block is the data delivered from the MAC layer to the physical layer.

Wi-Fi: The wireless channel access in IEEE 802.11 network is determined by Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) random access mechanism. Under CSMA/CA, when the channel is busy, a node does not transmit. When the channel is idle, a node has to wait for Inter Frame Space (IFS) time and then backoff transmission for random number of slots before it can attempt to transmit. If the channel is still idle, the node can transmit. An exponential back-off algorithm runs if a packet loss is detected. A node can transmit a Maximum Transmission Unit (MTU) of size 1500 bytes after gaining access to the channel. The packets are fragmented/defragmented to send the data of MTU size.

Physical Layer

5G NR: To simulate the physical layer, we have used an abstraction model for 5G-NR as suggested by 3GPP. We have adopted the Rural Macro (RMa) path loss model along with log-normal shadowing to simulate the radio link for 5G-NR [84]. Once we have determined the Signaling to Noise Ratio (SNR) for a link using RMa model, Channel Quality Index (CQI) is evaluated. In [85], 3GPP has suggested a map to determine the Modulation and Coding Scheme (MCS) and code rate based on the CQI for a link. The MCS value and the number of RBs allocated to a link are used to determine TBS. We have implemented the algorithm for evaluating TBS in the simulator as given in [85]. The different parameter values for calculating TBS are given in Table 4.1. A packet loss occurs whenever the MCS and code rate (based on measured SNR) are higher than the allowed MCS and code rate for the current SNR. Note that the SNR is measured and updated after every 10 ms.

Wi-Fi: The wireless channel for IEEE 802.11 network is modeled using the Urban Micro (UMi) path loss model described in [86]. We have also considered log-normal shadowing along with the path loss. We have adopted SNR to MCS/code rate map of a standard outdoor Wi-Fi device to evaluate the channel capacity. Cisco Aironet 1570 device is selected as it is suitable for outdoor deployment of Wi-Fi APs [87]. To simulate packet losses, we have used the SNR to Packet Error Rate (PER) map given in [88].

Table 4.1: Network Model

Parameters	**Values**
Number of Wi-Fi APs	10
Number of Users	80
Packet Payload	1000 bytes
TCP Header	60 bytes
5G-NR Network Model	
Carrier Frequency	1.9 GHz (TDD)
5G Numerology	1
Bandwidth	60 MHz (N_{RB} = 162 PRBs, 2 slots per sub-frame)
PDSCH Config	OFDM symbols = 14 (Normal CP), DMRS symbols = 1, No PTRS
gNB coverage radius	1000 m
UE Speed	Stationary
UE/gNB Transmit Power	23/30 dBm
UE/gNB Antenna Gain	2/6 dBi
UE/gNB Antenna Height	1.5/35 m
UE/gNB Noise Figure	7/10 dB
Wi-Fi Network Model	
Bandwidth	20 MHz
Operating Frequency	2.4 GHz
Coverage radius	40 m
UE/AP Transmit Power	15/20 dBm
AP Antenna Gain	4 dBi
AP Antenna Height	10 m
MPDU	1500 bytes

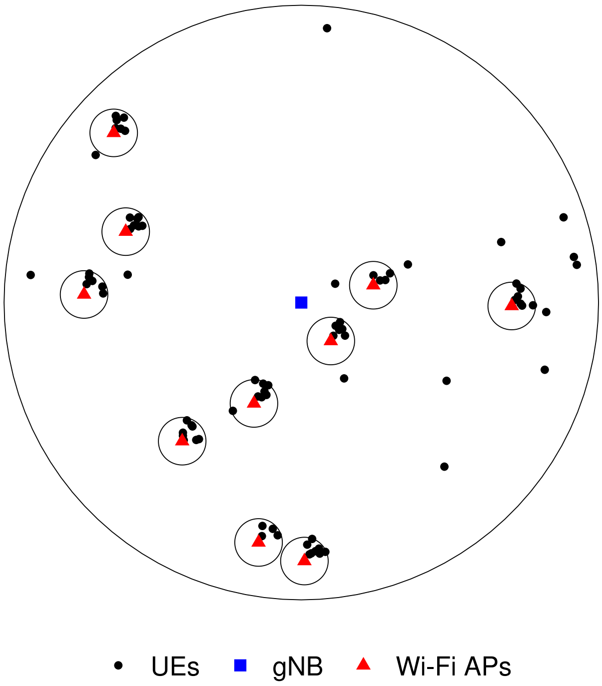


Figure 4.3: A single cell simulation scenario with 1000 m cell radius and multiple Wi-Fi basic service areas

4.2 Performance Analysis

We consider a single-cell scenario with 1000 m radius as shown in Figure 4.3. The network model comprises a 5G NR cell and multiple Wi-Fi Basic Service Areas (BSAs). The cell consists of a gNB-NR entity located at the center. We assume that a PoP is present near the vicinity of gNB-NR. The 5G-Flow controller and the MRN OpenFlow switch can either be co-located with the gNB-NR entity or can be cloud-based. Multiple Wi-Fi APs provide radio coverage (BSA) inside the gNB-NR cell and are distributed uniformly. A UE in the network is assumed to have two radio interfaces: 5G-NR and Wi-Fi. We also assume that 80% users are connected to both the RATs while the remaining users are connected to only gNB-NR. The users, in our simulation model, are assumed to be stationary. We perform Monte Carlo simulations for 1 second and average the results over 50 deployment scenarios. We analyze average TCP throughput and average packet delay for the 5G-Flow network and compare them with the current 3GPP 5G network.

Table 4.2: Simulation cases for 80 users in a cell requesting different data services

Cases	Service Priority (Streaming Service Type)			
	1 (Mission-critical)	2 (Real time)	3 (Non-real time)	4 (Best-effort)
Case a	0	0	40	40
Case b	10	10	30	30
Case c	20	20	20	20
Case d	30	30	10	10
Case e	40	40	0	0

4.2.1 Downlink dataflow management for UEs with different service types

As discussed before, data flow management involves selecting an optimal RAT (from the available RATs) for each UE session based on the RAN-level information. To analyze the performance of 5G-Flow RAN concerning RAT selection, we consider four types of data services offered in the network in the order of priority, i) mission-critical streaming, ii) real-time streaming, iii) non-real-time streaming, and iv) best-effort. We assume that a UE requests only one type of service for the entire duration of simulation for simplicity. We assume Poisson traffic with the packet arrival rate of 500 packets/sec. The packet size is fixed at 1000 bytes (payload). Therefore, the downlink bit-rate is 4 Mbps for all the users (irrespective of the service requested), and TCP protocol is used as the transport protocol. Since we aim to analyze the downlink performance, we assume that the entire available bandwidth is used by the downlink traffic and do not consider the bandwidth used by TCP-ACK packets (in uplink). We evaluate the uplink performance of the proposed network in the scenario discussed next. We assume 80 users in our simulation model. We consider five different cases based on the type of service requested by a user, and these are explained in Table 4.2. The 5G-NR radio resource scheduler uses a priority scheduling algorithm to give better service to higher priority users. Wi-Fi network uses CSMA/CA to schedule the users and does not distinguish between user service priorities.

For performance evaluation of downlink dataflow, we have designed a threshold-based heuristic algorithm for RAT selection given in Algorithm 1. The RAT-specific attributes and the metrics considered in the algorithm are provided via OpenFlow switches in the 5G-Flow network. As discussed before, the controller can subscribe to measurement data at the OpenFlow switch via NETCONF notifications. The algorithm considers the following metrics. The load conditions at RATs (Wi-Fi and gNB-NR) and 5G-NR radio channel quality are the most critical metrics to select a RAT for UE efficiently. Besides, the RAT selection algorithm considers the type (priority) of service being used. Due to the small coverage area of Wi-Fi AP, UEs' received signal strength does not vary much. Hence, the channel condition for a UE under its associated Wi-Fi network is always considered good. We assign different weights for the different metrics and calculate the value of T_i (for every UE) based on the following equation. We then select the RAT for a UE based on the predetermined threshold value. T_i can be evaluated as follows:

$$T_i(l_g, l_w, ch_{g,i}, s_i) = \alpha \cdot l_g + \beta \cdot l_w + \gamma \cdot ch_{g,i} + \delta \cdot s_i, \tag{4.3}$$

where,

$$\begin{aligned} l_g &= \text{Load at gNB}, \\ l_w &= \text{Load at Wi-Fi AP associated with user } i, \\ ch_{g,i} &= \text{Channel condition for user } i \text{ under 5G-NR network}, \\ s_i &= \text{Service type of user } i. \end{aligned}$$

Let $L = \{1, 2, 3\}$ denote the set of values representing low, medium, and high load respectively at a particular RAT. The load at gNB (l_g) and Wi-Fi (l_w) take its values from L. The channel condition experienced by user i from gNB-NR is represented by $ch_{g,i} \in \{0, 1\}$, where 0 represents good channel and 1 represents bad channel. We distinguish between good and bad channels based on a threshold value of received SNR, fixed at 6 dB in the simulation model. The service type of user i is represented by $s_i \in \{1, 2, 3, 4\}$, where 1 represents the highest priority service while 4 represents the service type with the least priority. The coefficients in Equation 4.3 (α, β, γ and δ) are the weights assigned to the metrics in the equation. The weights can be modulated based on the impact of a certain metric on system performance. The weights considered in our algorithm are given in Algorithm 1. We assign the highest importance to $ch_{g,i}$ because a bad channel leads to resource wastage and poor performance. If the gNB-NR RAT is observing high load, the high value of α ensures that additional users are not scheduled. This ensures that the QoS of already scheduled high priority users under gNB-NR is not impaired. After the evaluation of T_i, RAT selected for a user i (R_i) is determined based on the threshold value $T^{'}$. Since we aim to balance the load on the available RATs, $T^{'}$ is estimated to be the mean of all possible values of T_i. R_i is evaluated as follows:

$$R_i = \begin{cases} 1, & T_i > T^{'} \quad \text{(1 represents Wi-Fi)}, \\ 0, & T_i \leq T^{'} \quad \text{(0 represents gNB)}. \end{cases} \tag{4.4}$$

Since the 5GC is unaware of the RAN level information, the RAT selection policy for standard 3GPP 5G network only considers the service priority. For the performance evaluation of the 5G network, we consider that a user with service priority 1 or 2, is always scheduled at gNB, and with service priority 3 or 4 is always scheduled at Wi-Fi (irrespective of load or channel condition).

Results

Figure 4.4 shows the average TCP throughput and packet delay for various simulation cases considered in Table 4.2. Here, the *TCP throughput* is average transport layer throughput for all users under one service type. The *packet delay* is the end-to-end delay in the 5G-Flow system, between application layer at the multi-RAT network and at UE or vice-versa. We first analyse the graphs in Figure 4.4a and Figure 4.4e, for extremely skewed service type distribution among users. Since the 5G-Flow network considers load on RATs and channel conditions, RAT selection is implemented efficiently. Hence, the TCP throughput and delay performance for the 5G-Flow network are significantly better. Since all the users in standard 5G network are scheduled either on Wi-Fi RAT (in Figure 4.4a) or gNB-NR RAT (in Figure 4.4e), the respective RAT is overloaded, worsening the overall performance. This also leads to poor utilization of the other RAT.

Figure 4.4b and Figure 4.4d illustrate average throughput and packet delay results for service type distributions which are relatively less skewed than the cases discussed before. In Figure 4.4b, the average packet delay under standard 5G network is quite high for users with service priority 3 and 4 as all 60 users (with these service priorities) are scheduled at Wi-Fi. Under this case, gNB experiences low load, and hence the average

Algorithm 1: Downlink RAT Selection Algorithm

Input:

$C_g = C_0$ // gNB channel capacity

$C_w = W_0$ // Wi-Fi channel capacity

$l_g = 1, l_w = 1$ // Starting at low load

$\alpha = 30, \beta = 10, \gamma = 50, \delta = 25$

$T' = 170$

// $N =$ Number of UEs associated with both RATs

for i *from* 1 *to* N **do**

    Evaluate $ch_{g,i}$ for user i

    Calculate $T_i(l_g, l_w, ch_{g,i}, s_i)$ based on Equation 4.3

    if $T_i > T'$ **then**

        $R_i = 1$

        Decrease C_w

        if $C_w = medium \mid\mid C_w = low$ **then**

            Update l_w

        end

    else

        $R_i = 0$

        Decrease C_g

        if $C_g = medium \mid\mid C_g = low$ **then**

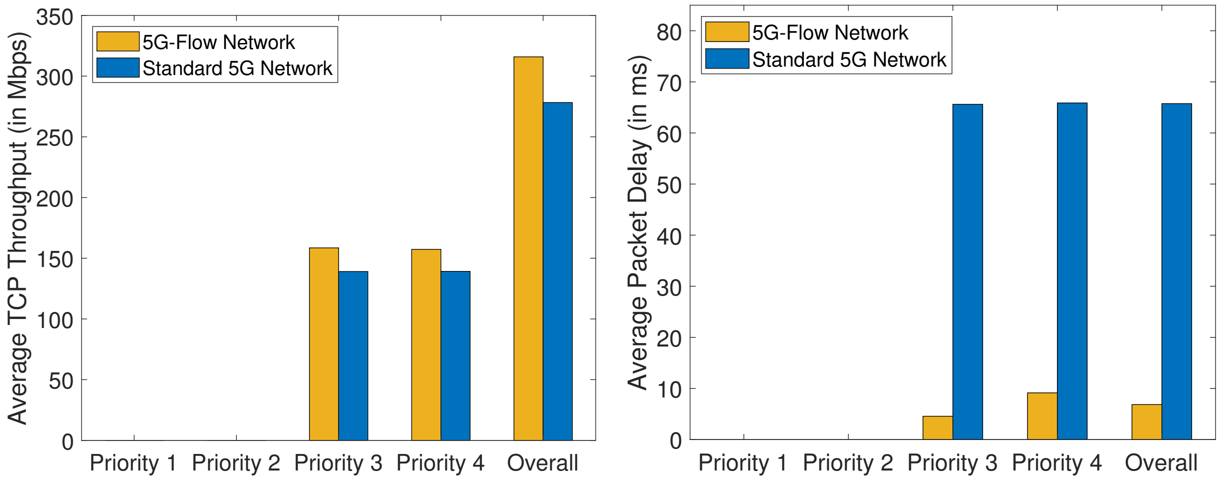
            Update l_g

        end

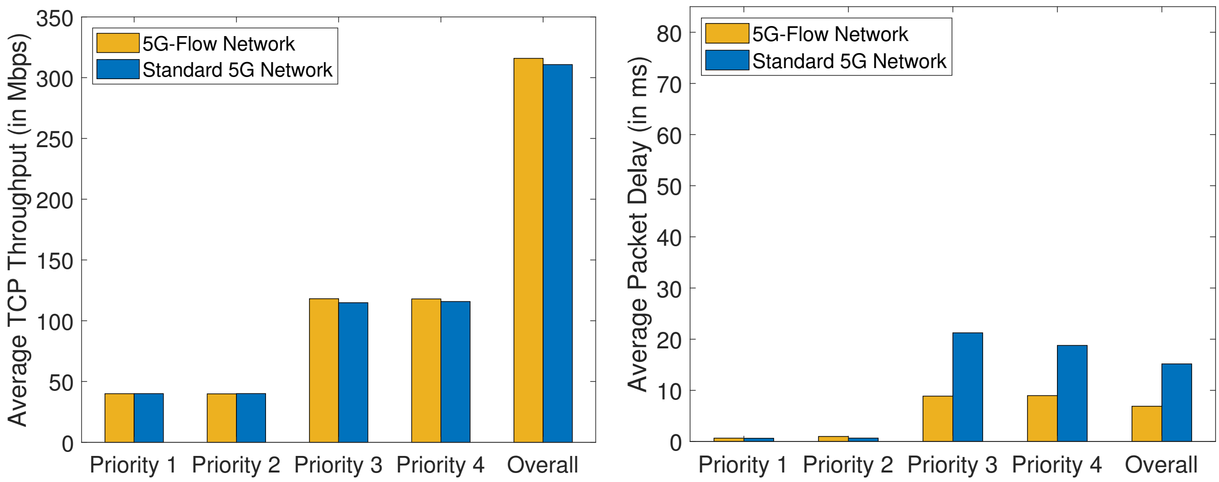
    end

end

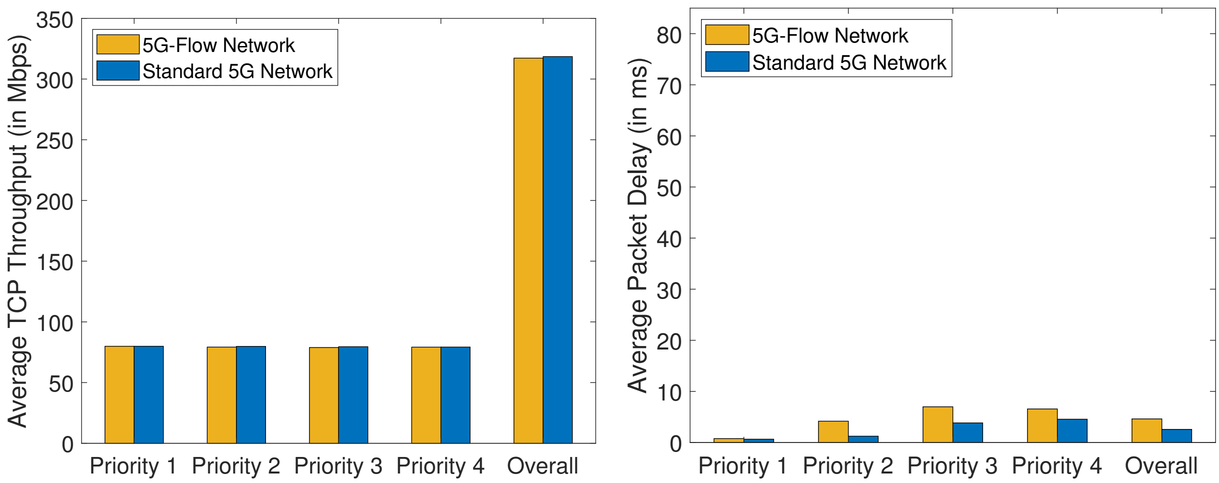
Output: $R_i, \forall i \in \{1, 2, \ldots, N\}$



(a) Case a: $\{0, 0, 40, 40\}$

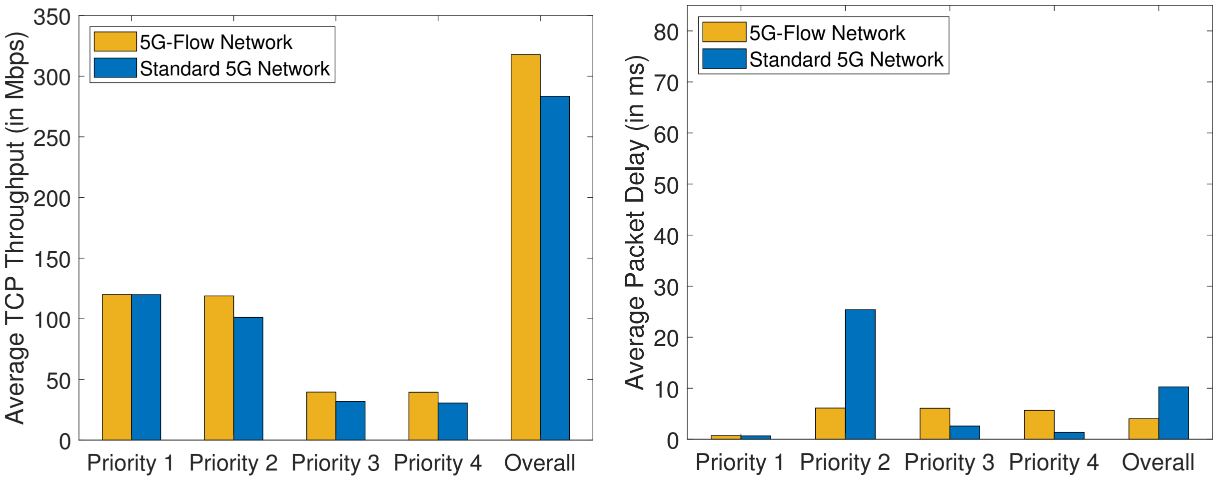


(b) Case b: $\{10, 10, 30, 30\}$

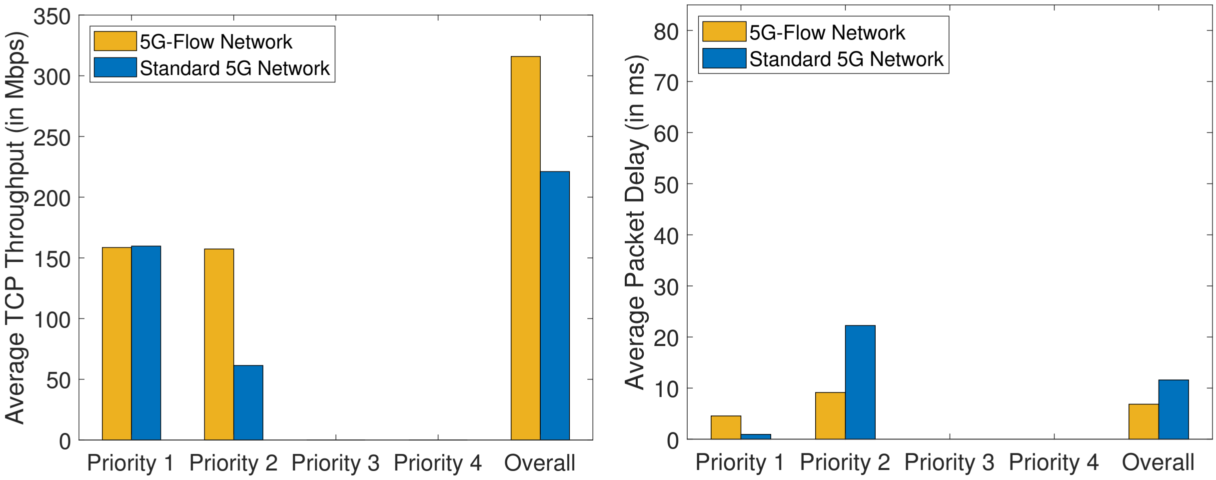


(c) Case c: $\{20, 20, 20, 20\}$

Figure 4.4: Performance evaluation of 5G-Flow network with respect to standard 5G network when users with different service priorities are present *(cont.)*



(d) Case d: $\{30, 30, 10, 10\}$



(e) Case e: $\{40, 40, 0, 0\}$

Figure 4.4: Performance evaluation of 5G-Flow network with respect to standard 5G network when users with different service priorities are present

packet delay is quite low for users with service priority 1 and 2. Similarly, in Figure 4.4d, the delay performance of users with service priority 2 suffers as the gNB experiences traffic load from 60 users, and it prioritizes the users with priority 1. The performance of evenly distributed user service priority is presented in Figure 4.4c. The performances of 5G-Flow network and standard 5G network are comparable as the 5G network follows a RAT selection policy that is best suited for evenly distributed user service priority.

In general, the standard 5G network cannot efficiently use the available radio resources in a multi-RAT scenario. The performance of the 5G-Flow network significantly improves as it considers RAN-level information while performing RAT selection.

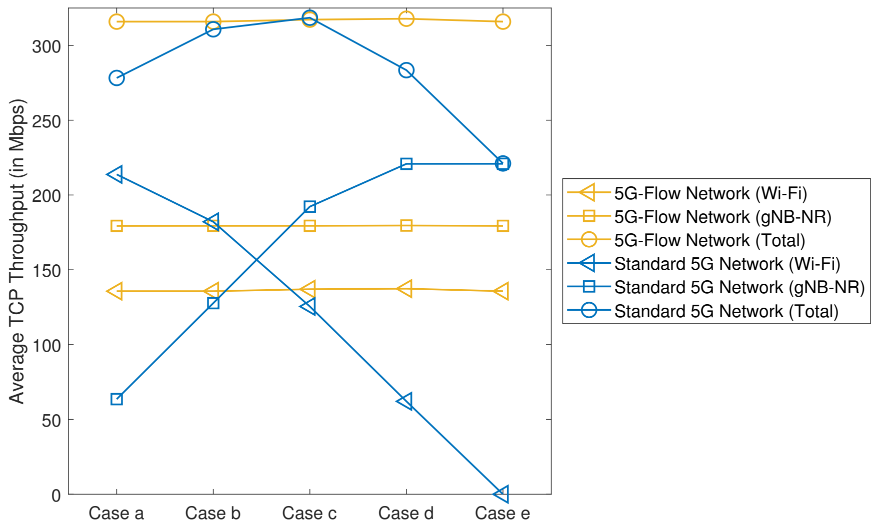


Figure 4.5: Load balancing across RATs

Load Balancing: In Figure 4.5, the graph shows how the traffic is split across various RATs for simulation cases presented in Table 4.2. The 5G network cannot balance the traffic across RATs; hence the overall throughput suffers. On the other hand, irrespective of considered cases (with diverse service requests from users), the 5G-Flow network can efficiently balance the load across available multi-RAT resources and deliver a consistent throughput performance.

4.2.2 Uplink Dataflow management

The proposed 5G-Flow network enables flexibility in managing uplink and downlink traffic of every UE independently. Moreover, our architecture decouples uplink and downlink traffic management, i.e., RAT selection for a UE's uplink and downlink traffic is determined independently. Due to this decoupling, we expect that the performance of the overall network and, in particular, of the Wi-Fi network will improve when the uplink users are fewer in a Wi-Fi network. This improvement is expected as fewer users will contend to gain access to the channel in a CSMA/CA system.

In the existing 3GPP 5G network, the ATSSS feature allows for independent decisions on uplink and downlink traffic distribution. UE selects RAT for uplink transmission based on channel quality observed on multiple RATs and ATSSS rules. UE does not know the load on each RAT to decide the best possible RAT for uplink. For instance, let us assume that a UE observes a good channel gain from Wi-Fi AP as well as gNB-NR, so it selects the gNB-NR interface for uplink. gNB-NR may be heavily loaded, and Wi-Fi may be lightly loaded, leading to poor uplink performance (which could have been better on Wi-Fi). Therefore, without loss of generality, we assume that UEs select the same RAT for uplink and downlink, i.e., uplink and downlink are essentially coupled for 5G network.

In our simulation scenario, we use TDD duplex scheme to allocate resources for uplink and downlink. The transmission periodicity for the TDD radio frame is 5 ms, i.e., the frame repeats after 10 slots (for 30 kHz sub-carrier spacing). TDD frame configuration is {D,D,D,S,U,U,U,U,U}, where D, U, and S represent downlink, uplink and special slot respectively [89]. The S slot consists of downlink symbols except for the last symbol, which is used as a switching symbol. We consider 80 users each having 3 Mbps downlink, and 1 Mbps uplink data rate requirement. We consider the Poisson traffic model for video streaming application. Since we do not consider service priority in the uplink scenario, we use a round-robin scheduler for 5G-NR to allocate radio resources to the users.

For those users that are dual-connected in the 5G-Flow network, downlink data is scheduled on Wi-Fi. However, we schedule the uplink data to gNB-NR interface for users who are experiencing good channel gain from the gNB. The RAT selection algorithm for uplink users is given in Algorithm 2.

Results

Here, we consider average TCP throughput for all the users in downlink direction and uplink direction separately. Figure 4.6 demonstrates that downlink TCP throughput is higher for the 5G-Flow network as compared to the standard 5G network. We also observe the percentage of served traffic to offered traffic for Wi-Fi APs in our simulations. While it is 75.87% under standard 3GPP 5G network, it is 88.5% for 5G-Flow network. This improvement in 5G-Flow network is observed due to the reduction in load at Wi-Fi APs as we have efficiently managed uplink users based on Algorithm 2. Therefore, fewer users contend to gain access to channel under Wi-Fi APs in 5G-Flow network, thereby increasing performance efficiency of Wi-Fi APs. For the same reason, the delay performance has also significantly improved.

Algorithm 2: RAT Selection Algorithm for Uplink

Input:

$C_g = C_0$ `// gNB-NR Uplink channel capacity`

`/*` $W_0 =$ `Maximum number of users allowed to connect with a Wi-Fi AP in uplink */`

$A =$ List of sorted Wi-Fi APs (in descending order) with respect to number of connected UEs

for *each* i *in* A **do**

  $N_i =$ Number of users connected to Wi-Fi AP i

  $K = \min(N_i - W_0, C_g)$

  Move K users (closest to gNB) from Wi-Fi to gNB-NR for uplink data transmission

  Update C_g

end

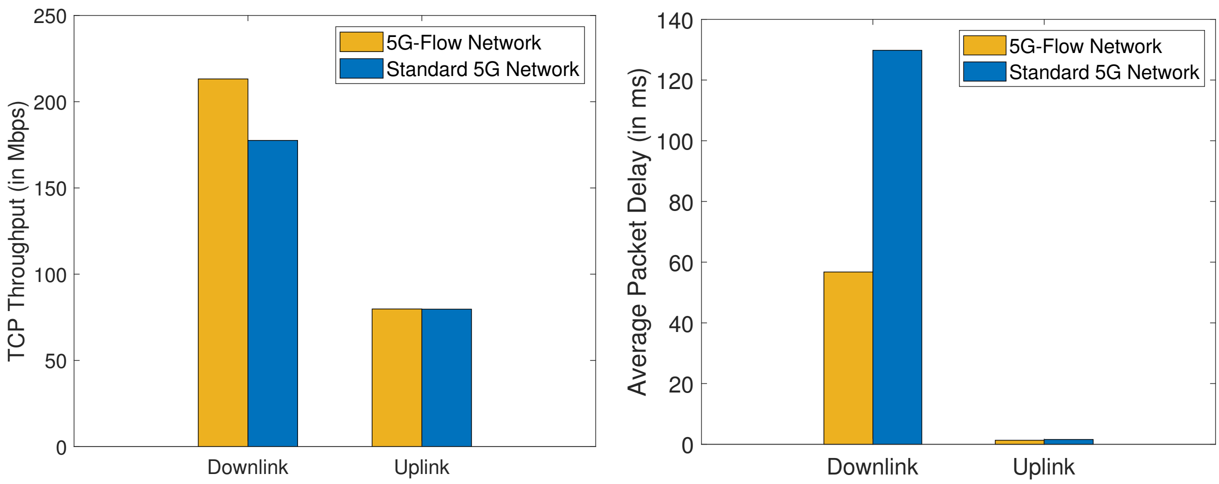


Figure 4.6: Performance comparison between 5G-Flow network and standard 5G network when uplink and downlink data flows are decoupled

4.3 Summary

In this chapter, we have presented a system-level simulator developed in MATLAB for the 5G multi-RAT network. We have discussed the simulation steps and network stack implementation. Using this simulator, we have analyzed the performance of data flow management across multiple RATs in a 5G-Flow network. We have compared the performance of the 5G-Flow network with the standard 5G network. Results for downlink data flow management for UEs requesting different service types have shown that the 5G-Flow network uniformly balances load across available RATs, efficiently using the multi-RAT resources. The performance evaluation of uplink data flow management has shown significant improvement in the Wi-Fi network performance under the 5G-Flow network due to the flexibility of uplink/downlink decoupling, which is not present in the 3GPP 5G network.

Chapter 5

Efficient Middle Mile Network Planning

The Frugal 5G network architecture proposes to use a wireless middle mile network to backhaul Wi-Fi APs in the villages. The proposed network architecture also integrates the middle mile network under the unified multi-RAT framework that ensures QoS guarantees to the end-users. In this chapter, we center our focus on a crucial aspect of the middle mile network, i.e., planning its topology. The middle mile network comprises *middle mile clients* and *middle mile AP*, wherein AP is co-located with PoP and clients are deployed in the village clusters. As discussed in Chapter 1 (Section 1.4), the rural areas are sparsely populated with large empty spaces between the clusters. Therefore, to extend connectivity from PoP to the village clusters, we consider a fixed wireless multi-hop network topology for the middle mile network. A multi-hop network is also essential for overcoming obstructions such as hills by creating wireless hop over it. Now, to create a multi-hop path between the AP and a client (located far off from each other), we introduce a new element in the middle mile network — an *Intermediate Node (IN).* The INs are beneficial in extending the network coverage and enhancing the network capacity cost-effectively. An IN is similar to a middle mile client, but it does not directly serve the Wi-Fi APs. It only decodes and forwards the data from one middle mile node to another.

The key objective of planning the middle mile network is to optimize the network performance while minimizing the infrastructure requirement, thereby proposing a cost-effective solution. Towards this, we first articulate the problem. It can be stated as: *Given the location of village clusters and PoP, determine the network topology that maximizes network throughput while minimizing the number of hops and INs in the network.* We minimize the number of hops to minimize the delay in the network, thereby enhancing the network performance. To reduce the infrastructure cost, the number of INs in a multi-hop network should be minimized. Since the proposed problem has high complexity, we propose a heuristic algorithm as a solution. We model the middle mile network as a graph and consider a set of potential IN locations in the graph. We then solve an IN selection problem with a vast search space using Simulated Annealing (SA). SA is a stochastic search-based algorithm for approximating the global optimum. We also compare the performance of the proposed algorithm with a greedy approach to solving the problem.

We begin by discussing the prior research works related to planning of wireless multi-hop networks (Section 5.1). We then present the system model and problem formulation (Section 5.2). We propose a simulated annealing based heuristic algorithm to plan the middle mile network (Section 5.3). The performance analysis of the algorithm is presented next (Section 5.4). We summarize the chapter in the last section (Section 5.5).

5.1 Prior Work

In this section, we discuss the relevant research proposals for the planning of multi-hop networks incorporating relays (similar to INs). In multi-hop networks, relay deployment is a desirable solution to enhance connectivity and network performance. Most of the works regarding relay placement have been proposed for ad-hoc networks such as wireless sensor networks. Relay placement in infrastructure-based networks has gained popularity recently as it provides a low-cost backhaul alternative.

In wireless sensor networks, the key objectives of relay placement strategies are network lifetime enhancement, fault tolerance, and connectivity [90–93]. On the other hand, for infrastructure-based networks, throughput and coverage are major concerns in addition to connectivity. Also, the network is dynamic in wireless sensor networks as compared to static in infrastructure-based networks. Therefore, the strategies studied in wireless sensor networks can not be directly applied in infrastructure-based wireless networks such as WiMAX, LTE Advanced, wireless mesh networks, and the proposed multi-hop middle mile network. Therefore, we study the research developments on relay placement in infrastructure-based wireless networks.

In [94], Lu et al. develop a network planning model for an IEEE 802.16j relay network. Optimal placement of base station and relay nodes is considered where candidate site, user demand, and cost information are given. The authors solve this problem by solving an integer programming problem for a two-hop scenario. However, the authors do not consider link rate allocation and user data rate requirements while designing the model. In [95], Wang et al. aim to maximize the system capacity in a two-hop network by placing the relay optimally. The relay selection rules are based either on signal strength or throughput. They restrict a relay's position to be on a circle around the base station and solve a single variable optimization problem. In [96], Abichar et al. formulate a mixed-integer linear programming problem to find the number and locations of relay nodes for a given customer base. However, they only consider a two-hop scenario for the given WiMAX network.

Several other schemes, presenting budget constraint relay placement in the network, have been discussed in the past. In [97, 98], Lu et al. propose a joint deployment scheme for BS and relays. The authors formulate an integer linear programming problem to maximize the system capacity under a given budget. The authors propose a two-stage network deployment algorithm, wherein the first stage involves deployment of BSs and in the second stage relays are deployed to serve the users not served by BSs. Also, they consider only two-hop relaying in their problem formulation to limit complexity. In [99], Chang et al. also aim to maximize the system capacity under budget constraint in a network of one BS and k relay stations. This is achieved by placing the relays optimally. They partition the area into k sub-areas and calculate the best location in each sub-area. This may lead to unnecessary deployment costs as a relay per sub-area might be more than required in regions with lower user density. The authors extend their work in [100] and aim to minimize the number of relays deployed while enhancing the throughput performance of the network. However, their work is limited to a two-hop scenario.

Prior work has not focused on more than two hops to limit complexity in a multi-hop infrastructure-based network. As a result, minimizing the hops and minimizing the

INs/relays in a multi-hop scenario has not been studied. This is a critical requirement in rural areas where we have to serve a large area with limited resources.

5.2 System Model

We consider a geographic area in a rural setting that needs to be served via a Frugal 5G network. We deploy WLAN APs in these clusters to serve the end-users. These clusters are, in general, uniformly scattered over the given area. The middle mile network enables the connectivity between the WLAN APs and the PoP. We assume that a middle mile AP is located at the center of the given area and has a wired backhaul via the PoP. We further assume that there is one middle mile client situated in each village cluster. Since the middle mile clients are located far away from each other, it is difficult to connect them directly. Thus, we employ INs to connect them and form a wireless multi-hop middle mile network. The middle mile clients and INs are identical multi-hop nodes but differ in the fact that middle mile clients backhaul the WLAN APs, while the INs do not. An IN can be easily used to serve a WLAN AP if demand arises in the future. The middle mile clients (in clusters) aggregate the data from the WLAN APs and transport it to the middle mile AP (near PoP), with the help of INs. An instance of a middle mile network is depicted in Figure 5.1. A single middle mile client may serve many WLAN APs, and the connectivity between them may be wired or wireless.

We model the middle mile network planning as a graph-theoretic problem. The network is modeled as an undirected graph $G(V, E)$ where V is the set of nodes in the network and E is the set of edges between the nodes. The set V comprises the middle mile AP, the middle mile clients, and the INs used to connect them. The set of middle mile clients is denoted by $M = \{1, 2, 3, \ldots\}$ and $|M|$ denotes cardinality of set M . The set of INs is denoted by $L = \{1, 2, 3, \ldots\}$ and $|L|$ is the cardinality of set L. The number of hops between the middle mile AP and the middle mile clients M is denoted by $H = \{h_1, h_2, h_3, \ldots, h_M \mid h_i \in \mathbb{N}\}$. Assume that the maximum range of an IN or a middle mile client to be fixed and is denoted by r m. A wireless link between any two nodes of a middle mile network is assumed to be a point-to-point link.

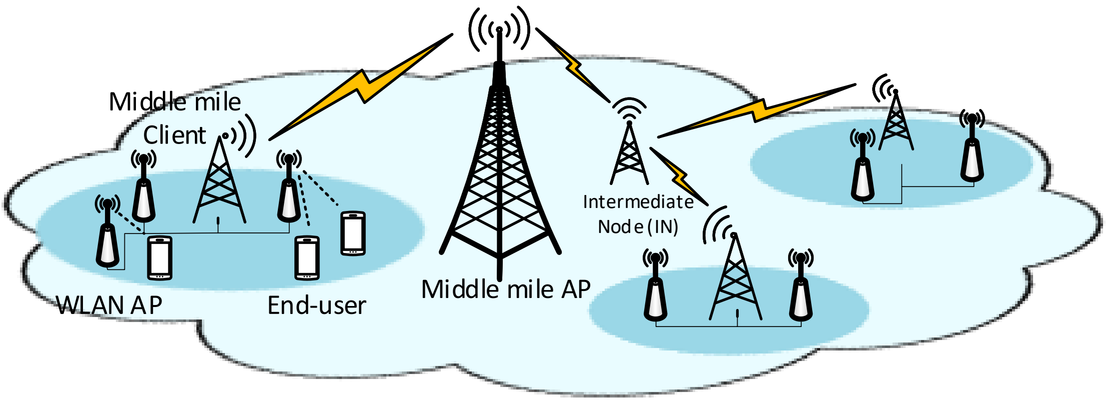


Figure 5.1: An instance of the middle mile network

5.2.1 Problem Formulation

Our objective is to maximize the network throughput and minimize the hops and the number of INs for optimally planning the middle mile network. The downlink throughput experienced by a middle mile client m is denoted by U_m. The network throughput is defined as the sum throughput of $|M|$ middle mile clients in the network graph $G(V, E)$ i.e $\sum_{m=1}^{|M|} U_m$. Mathematically, the problem can be written as:

$$G^*(V^*, E^*) = \underset{G(V,E)}{\arg\max} \sum_{m=1}^{|M|} U_m - f(H) - g(L). \tag{5.1}$$

Here, f and g are exponential penalty functions of H and L respectively. Exponential penalty functions are chosen to assign a high cost if the number of hops and INs is large. The function f is an average of hops between middle mile client and AP and is defined as $f(H) = \frac{1}{M}\sum_{m=1}^{|M|} e^{a_0 h_m}$. The function g is defined as $g(L) = e^{\frac{b_0|L|}{|M|}}$. The constants a_0 and b_0 in the functions are chosen to match the penalties closely.

Minimizing the INs/relays in a network is proven to be NP-Complete problem [101]. The considered problem not only aims to minimize INs but also minimizes the hops in the network along with maximizing the throughput. Therefore, to solve this problem efficiently, we propose a low-complexity heuristic algorithm in the next section.

5.3 Middle Mile Network Planning

To develop a heuristic algorithm for the proposed problem in Section 5.2.1, we model it as an IN selection problem. We consider a set of potential IN locations in a given area.

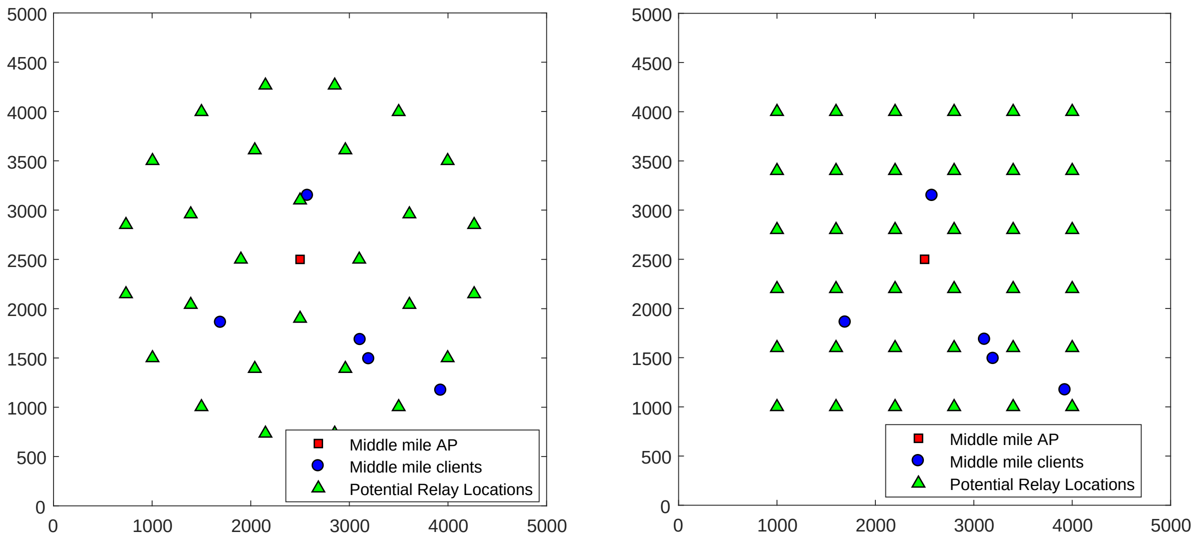
subsequently, we run a selection algorithm over these potential IN locations to attain our objective. The potential IN locations are chosen strategically instead of placing them randomly. The arrangement of these IN locations is described next.

5.3.1 Arrangement of Potential Intermediate Nodes

We present two possible arrangements of potential IN locations on which we run our selection algorithm. Figure 5.2a shows a circular structure of potential IN locations and a set of middle mile clients in a given area. A circular grid is chosen because the contours of constant received power are circular in nature. A square grid may also be considered, as shown in Figure 5.2b, for its regular structure. Next, we need to determine the best acyclic multi-hop graph that spans over the selected INs and the middle mile nodes.

5.3.2 Intermediate Node Selection

We present a SA based algorithm to solve the IN selection problem. SA, a Markov chain Monte Carlo method, is a meta-heuristic search-based technique to estimate the global optimum of a given function. When the search space is vast, such as in our case, SA can be used to estimate the optimal solution. SA is inspired by annealing in metallurgy



(a) Circular grid for deploying INs

(b) Square grid for deploying INs

Figure 5.2: Potential IN placement locations (distances are in meters)

which involves heating and controlled cooling of a material. SA comprises two stochastic processes, i.e., i) generation of a new solution and ii) acceptance of a new solution. A temperature T controls the acceptance of any solution. It is important to note that we do not necessarily require a polynomial-time algorithm as the wireless backhaul planning is done only once before deployment or when a new node is added to the network. However, frequent re-planning with differing sets of conditions would require several time-consuming exhaustive search operations, thus necessitating an efficient solution to this problem.

We construct a Discrete Time Markov Chain (DTMC) on the states of the problem under consideration. Then, we transition on these states to attain the global optimum. A DTMC is characterized by its states, how neighboring states are constructed, and the Transition Probability Matrix (TPM). We explain these elements next.

States

The state refers to the different feasible graphs G_i that can be formed using the middle mile AP, middle mile clients, and INs. A feasible graph is a multi-hop tree with middle mile AP as the root node and only middle mile clients as the leaf nodes. Each state, i.e., a feasible graph G_i, has an associated weight $W(G_i)$. Let us denote G^* as the global optimum, i.e., the state with maximum weight $W(G^*)$. We define weight function as the objective of our optimization problem (Section 5.2.1) and it can be stated as $W(G) := \sum_{m=1}^{M} U_m - f(h_m) - g(L)$.

Neighboring States

We now explain how the neighboring states are determined in the DTMC. In our scenario, a neighbor graph is generated by selecting one of the two actions, i.e., *adding an edge* or *removing an edge*. Through this process, the SA algorithm can traverse through all the feasible graphs. Assume a feasible graph G_i. A graph G_j is its neighbor if obtained by choosing one of the following two actions — A_1 or A_2.

- *Adding an edge (A_1):* When an edge is added to a graph, a cycle may be formed, which renders the graph infeasible. Therefore, we randomly remove an edge such that the cycle is removed and no middle mile client is disconnected.

- *Removing an edge (A_2):* When an edge is removed from a graph, middle mile clients may be disconnected, making the graph invalid. Therefore, we can only remove those edges which do not disconnect the middle mile clients.

Transition Probability Matrix

Now, we define the TPM for the DTMC constructed above. Let p_{ij} be the transition probability from state G_i to its neighboring state G_j. The transition happens in the following two steps.

- In the first step, we choose an action, i.e., adding or removing an edge with equal probability i.e. $p^{A_1} = p^{A_2} = \frac{1}{2}$. Each action is executed in the following manner:

  - *Add:* Assume that e_i number of edges can be added to the graph G_i. The addition of an edge creates a cycle in the graph. If $l_i(\geq 1)$ edges can be removed from the graph to make it free from cycles, then the probability with which an edge can be added is $\alpha_i^{A_1} = \frac{1}{e_i}\frac{1}{l_i}$. The state remains unchanged if the edge, that is added to the graph, is removed.

  - *Remove:* Assume that k_i edges can be removed from the graph G_i. If $k_i = 0$, then state remains unchanged. The probability of action A_2 can be written as $\alpha_i^{A_2} = \frac{1}{k_i}$.

- In the second step, the weight of both the graphs, G_i and G_j is compared. If $W(G_j) \geq W(G_i)$, then transition happens with probability 1. If the $W(G_j) < W(G_i)$, then transition happens with probability $e^{(W(G_j)-W(G_i))/T}$, where T is the temperature parameter in the SA algorithm. The probability of this step can be written as $\beta = \min(1, e^{(W(G_j)-W(G_i))/T})$.

Now, the transition probability can be written as,

$$p_{ij} = \begin{cases} p^{A_1}\alpha_i^{A_2}\beta, & \text{Action 1,} \\ p^{A_2}\alpha_i^{A_2}\beta, & \text{Action 2,} \\ 0, & \text{when } G_j \text{ is not a neighbor of } G_i. \end{cases} \tag{5.2}$$

In the SA algorithm, we simulate the DTMC with the above transition probabilities as explained in Algorithm 3. These transition probabilities can be calculated in polynomial time and need not be stored. This ensures the computational viability of our approach. Next, we prove certain important properties of the constructed DTMC which ensures that there exists a unique steady state distribution.

Algorithm 3: Middle Mile Network Planning Algorithm

Input: $G_i \leftarrow$ Initial solution

$T \leftarrow T_0$
Calculate $W(G_i)$
while *Global loop condition not satisfied* **do**
    while *Local loop condition not satisfied* **do**
        $G_j \leftarrow$ Generate a neighbor graph from G_i
        if $W(G_j) > W(G_i)$ **then**
            $G_i \leftarrow G_j$
        end
        else
            $r \leftarrow$ Uniform random number between 0 and 1
            $p \leftarrow e^{(W(G_j)-W(G_i))/T}$
            if $r < p$ **then**
                $G_i \leftarrow G_j$
            end
        end
    end
    $T \leftarrow \alpha T$
end

Output: Result: G^*

Theorem 1. *The constructed DTMC $\{G_n, n = 0, 1, 2, \ldots\}$ is finite, aperiodic and irreducible.*

Proof. Assume that L_0 nodes are selected from L potential nodes. Hence, we have a total of $M + L_0$ nodes in the network. The total possible edges in the network are $E_0 = \binom{M+L_0}{2}$. The number of all possible graphs that can be formed using these edges is 2^{E_0}. Since the total feasible graphs are a subset of all possible graphs, the number of feasible graphs is finite. The DTMC has self-loops as there is a non-zero probability of remaining in the same state. Therefore, the DTMC is aperiodic. There is a path to go from any graph to any other graph by dropping all the edges and adding only the required edges. This proves that DTMC is irreducible. □

5.4 Performance Analysis

5.4.1 Simulation Framework

We consider an area of 5000×5000 m^2 in which middle mile clients are deployed uniformly at random. Without loss of generality, we assume that the middle mile AP is located at the center of the given geographic area. We consider the operational frequency to be 3.3 GHz and the bandwidth to be 40 MHz. The transmit power for all the nodes in the network is kept low at 20 dBm. We vary the number of middle mile clients in the given area from 5 to 10. In the simulations for the proposed algorithm, $a_0 = 2$ and $b_0 = 3$. The results are averaged over 100 different distributions for a given number of middle mile clients. Other simulation related parameters are given in 5.1.

Since we assume a fixed wireless network, we only consider the path loss while formulating the problem. The wireless channel effects such as fading and shadowing are not considered, however we consider a fade margin while calculating path loss. We employ Friis free space path loss model to calculate the attenuation of the signal over a link. The received power can be calculated as:

$$P_r = P_t + G_t + G_r + 20\log_{10}\left(\frac{c}{4\pi d f_c}\right) - N_0 - NF - FM. \tag{5.3}$$

Here, P_r is the received signal strength (in dBm), P_t is the transmitted signal strength (in dBm), G_t, G_r are transmitter and receiver antenna gains (in dBi), d is the distance between the transmitted and received antenna (in meters), f_c is the radio frequency used (in Hz) and c is the speed of light (in m/s).

We assume minimum SNR for successful detection of signal to be 10 dB. Based on the minimum SNR required, we evaluate the range of middle mile client and IN using Equation 5.3. The range is approximately 800 m based on the parameters used in Table 5.1. Depending on this range, we design the circular grid and square grid.

Table 5.1: Network Model

Parameters	Values
Bandwidth (B)	40 MHz
Center Frequency (f_c)	3.3 GHz
Transmit Power (P_t)	20 dBm
Noise Level (N_0)	-100 dBm
Transmitter Antenna Gain (G_t)	4 dBi
Receiver Antenna Gain (G_r)	4 dBi
Noise Figure (NF)	5 dB
Fade Margin (FM)	10 dB

The wireless link capacity can be modeled using the standard Additive White Gaussian Noise (AWGN) channel capacity given by:

$$C = B \log_2 (1 + P_r/N_0) \,. \tag{5.4}$$

Here, C is the capacity in bit/sec, B is the channel bandwidth in Hz, and P_r/N_0 is the received SNR. For the purposes of fading channels, we may assume that this can be converted to an outage-based rate computation. Here, the discussion is not specific to a particular wireless standard or technology. The proposed algorithm can be easily modified to suit any technology. Therefore, we consider a generalized scenario here.

Before probing into the results, we describe two potential approaches that can be used as a benchmark for performance comparison.

Exhaustive Search for IN Placement

Assume that we have L possible locations for placement of the INs. To determine the best INs, we have to search for all the possible combinations of INs which are 2^L. After selecting INs, we determine the best possible graphs formed from these INs. It can be observed that the computational complexity is exponential for the possible IN locations. Hence, an exhaustive search cannot be used to compare with the proposed algorithm.

Greedy Placement of INs

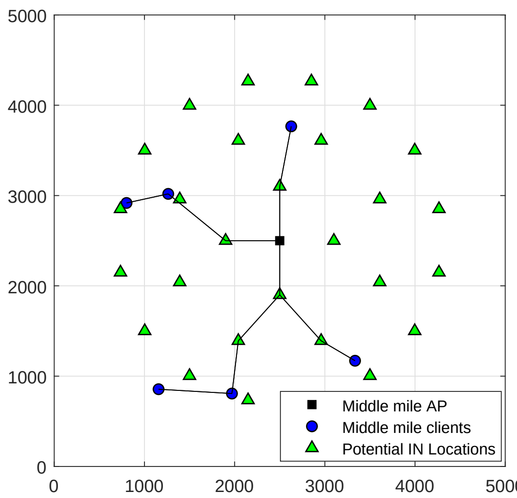
A heuristic method based on a greedy approach can be used to deploy INs in a network, and it serves as a benchmark to evaluate the performance of our algorithm. This is an iterative method in which we add an IN in every iteration. Here, we divide the given area into a grid. An IN can be placed at the center of any square in a grid. After selecting a position to place IN, we form a multi-hop graph using a *minimum spanning tree* algorithm. The graph that maximizes the weight $W(G)$ as defined above is chosen among all possible locations on the grid. The corresponding IN location is fixed. In the same manner, we deploy the INs iteratively.

5.4.2 Results

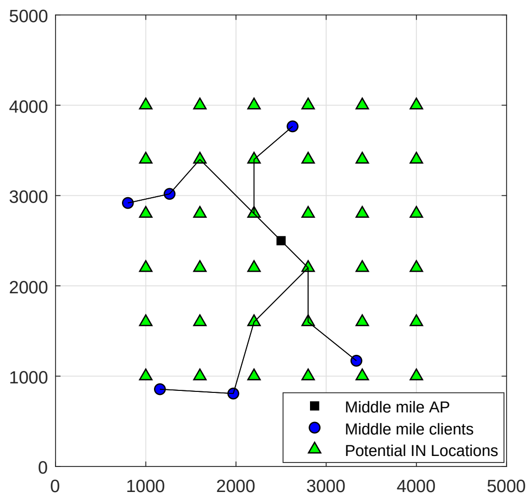
We now discuss the results obtained from our simulations in MATLAB [102]. As the computational complexity of the exhaustive search approach is exponential, we only compare our algorithm with the greedy approach. Figure 5.3 illustrates the deployment results for an example scenario obtained from our algorithm as well as the greedy approach. As proposed by our algorithm, the network topology includes uniformly chosen INs over the given area, improving the worst-case performance in the network. On the other hand, the greedy algorithm maximizes the throughput by making shorter, high-capacity links and compromises the throughput of other links.

Figures. 5.4a and 5.4b show the performance comparison of our proposed algorithm for a circular grid, square grid, and the greedy IN placement algorithm. There is a 24% improvement in network throughput of the circular grid based approach compared to the greedy IN placement. The square grid based approach shows 9% improvement in network throughput as compared to the greedy approach. Similarly, for the minimum user throughput, an improvement of 25% is observed for the circular grid compared to the greedy approach. These observations clearly state that a circular grid based approach not only maximizes the global outcome but also maximizes the individual user performance. We observe a saturation in the network throughput and a decline in minimum user throughput as the number of middle mile clients are increased. This is due to the increased number of branches in the multi-hop network limiting the offered capacity at each middle mile client.

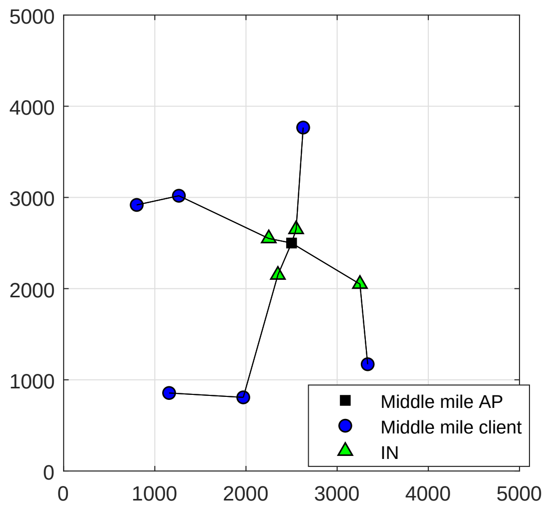
Figures. 5.4c, 5.4d and 5.4e show the comparison of three other parameters in the network. First, we compare the number of INs used for a given distribution of middle mile clients. A similar performance is observed for all three cases. This ensures that SA based technique is not increasing the infrastructure used to improve the performance. We then compare the average number of hops for all the middle mile clients in a given network. Also, the maximum number of hops among all the middle mile clients is plotted. It can be clearly observed that though the average number of hops for greedy is almost the same as the other two approaches, the maximum number of hops is the highest. The reason for this outcome is that the greedy IN placement does not have a global view. It favors



(a) Circular Grid ($\sum_{i=1}^{M} U_i = 827$ Mbps and $min(U_i) = 95$ Mbps)

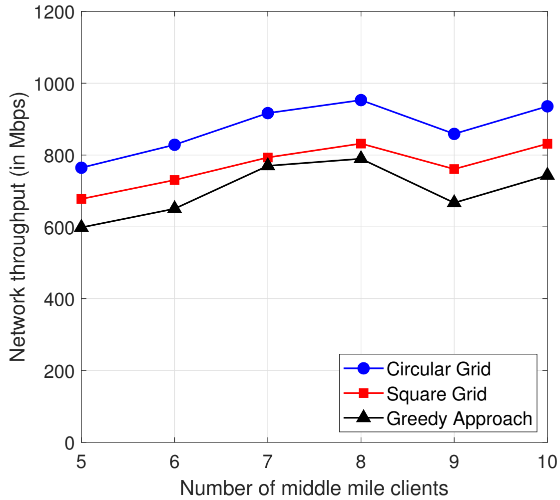


(b) Square Grid ($\sum_{i=1}^{M} U_i = 654$ Mbps and $min(U_i) = 109$ Mbps)

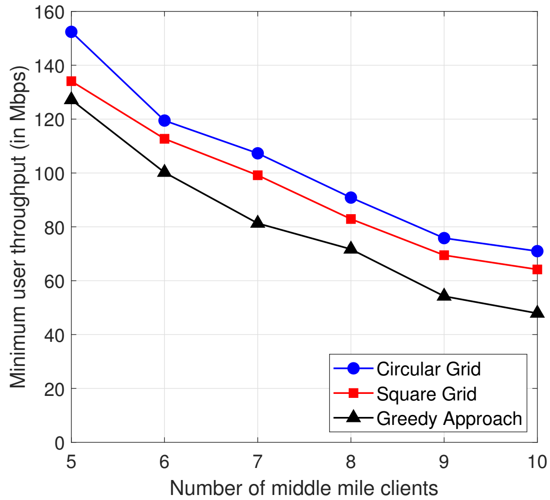


(c) Greedy IN Placement ($\sum_{i=1}^{M} U_i = 601$ Mbps and $min(U_i) = 80$ Mbps)

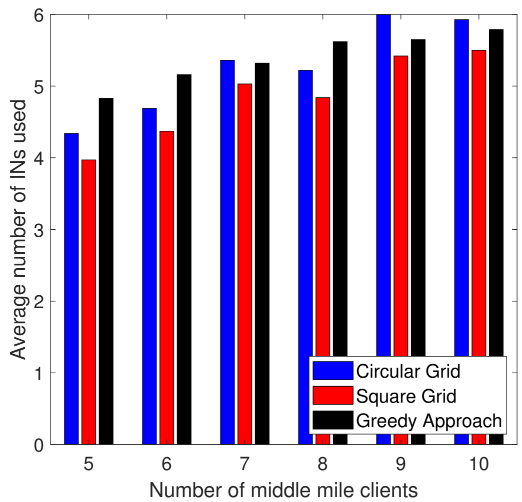
Figure 5.3: Comparison of a deployment example between the proposed algorithm for circular and square grid and greedy approach (distances are in meters)



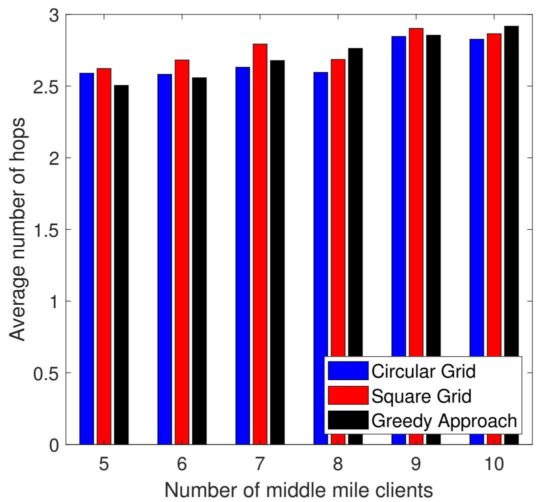
(a) Network throughput



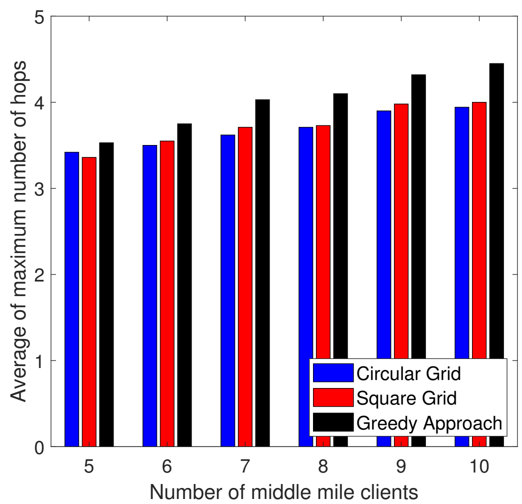
(b) Minimum user throughput



(c) Average number of INs



(d) Average number of hops



(e) Average of maximum number of hops

Figure 5.4: Comparative analysis of the proposed algorithm for circular and square grid, and the greedy IN placement algorithm

some middle mile clients while the performance for other middle mile clients suffers. Since the contours of constant power are circular, the circular grid is a much more practical alternative than the square grid. The observed performance of the circular grid based approach supports this conclusion.

The middle mile network design proposed here incorporates several design parameters for the Frugal 5G middle mile network, such as limited mobility support, large operational area, and relaxed network availability. It is observed that with incorporating 4-5 INs and allowing at most four hops, the performance of the middle mile network is significantly improved with our efficient planning algorithm. These INs are envisioned to be low power nodes and can cost-effectively operate on renewable energy. The proposed design is an excellent candidate for the middle mile network in rural areas.

5.5 Summary

In this chapter, we have focused on an important aspect of Frugal 5G network, i.e., planning the middle mile network. We have considered a fixed wireless multi-hop middle mile network. We have also introduced INs in the network to efficiently plan the network. We have formulated an optimization problem that aims to maximize network throughput while minimizing the number of hops and INs in the network. Since the proposed problem has exponential complexity, we have presented a low-complexity SA-based algorithm. The results have revealed that the proposed method performs much better than other approaches, including greedy design, even in the critical parameters such as minimum user throughput and maximum hop count. This can be extended to account for temporal and seasonal variations in the usage pattern to allow for higher redundancy with reasonable energy efficiency.

Chapter 6

Spectrum Sharing for Multi-Operator Middle Mile Network

In Chapter 2, we have discussed various desired characteristics of the Frugal 5G network, among which a key characteristic is *multi-operator resource sharing*. Resource sharing involves multiple operators sharing passive infrastructure (such as cell site, tower, or shelter), active infrastructure (such as antennas, base station, or backhaul), or spectrum. Spectrum costs are relatively high and form a large part of CAPEX for mobile networks. Therefore, efficient spectrum sharing leads to better spectrum utilization as well as significant reduction in CAPEX. This is aligned with the requirements of rural areas. To this end, we discuss spectrum sharing among multiple middle mile networks deployed by the same or different operators in this chapter[1]. We do not discuss spectrum sharing among macro BSs, but the proposed work can be extended to include macro BSs.

We consider several LTE-based[2] middle mile networks owned by multiple operators in a given area. The middle mile AP is an LTE-based eNB while the middle mile clients are LTE-based Customer Premises Equipments (CPEs). A CPE backhauls IEEE 802.11 based Wi-Fi APs, which in turn serve the end-user. Since our focus is not on middle mile

[1]The work presented in this chapter is done in collaboration with Sweety Suman.

[2]*Remark*: This work was completed in 2017, and at that time, the research related to the 5G cellular standard was in its early stage. Therefore, in this work, we have considered an LTE-based middle mile network.

network design, we consider a single hop topology for simplicity. We assume that a set of channels in the TV UHF band is available for rural areas. As discussed in Chapter 1, the TV UHF band (470-585 MHz) is highly underutilized in India. Additionally, lower frequency bands such as TV UHF bands can increase the middle-mile network coverage. However, when multiple operators coexist in the same TV UHF band, there is interference among middle mile networks leading to low spectral efficiency. The channels must be optimally shared among the operators to enhance spectral efficiency. It is equally essential to maintain fairness among operators while sharing the spectrum. Another crucial challenge concerning the proposed problem is that it should be based on trivial information that the operators can easily provide, considering they are reluctant to share sensitive network information. Towards this, we design a graph-theoretic algorithm for spectrum sharing among multiple operators. We employ Listen Before Talk (LBT) and Carrier Aggregation (CA) features provided by LTE standard while designing the algorithm [103, 104]. We evaluate the performance of the proposed algorithm via system-level simulations using ns-3.

The rest of the chapter is organized as follows. We start by discussing the prior work related to the proposed spectrum sharing problem (Section 6.1). We then discuss the system model (Section 6.2). Next, we present the spectrum sharing algorithm (Section 6.3). Further, we analyze the performance of the proposed algorithm through system-level simulations in ns-3 (Section 6.4). Lastly, we summarize the work presented in this chapter (Section 6.5).

6.1 Prior Work

The spectrum sharing problem discussed in this chapter is quite similar to the *limited spectrum pooling*, wherein a limited number of operators share a common pool of spectrum, obeying the rules of spectrum access set by multi-lateral agreements. Limited spectrum pooling has been studied in the context of heterogeneous networks in [105] and [106]. In [105], two operators pool equal bandwidth for sharing among small cells. An operator has preemptive priority over its share of pooled spectrum. This work mandates the sharing of scheduling information that operators might be reluctant to share. The operators report their spectrum occupancy information and interference level to the spectrum controller. Using the graph-theoretic or clustering approach, the spectrum controller allocates resources from the pool to each operator's small cells. In [106], spectrum sharing is studied for an LTE-based dense small cell network of two operators. In our system, the operators have equal priority over the spectrum in contrast to the pre-emptive nature of the scheme in [105]. Moreover, in [105, 106], scheduling information has to be shared among operators, which is infeasible in general.

In addition to the above works, *graph-theoretic techniques* for multi-operator coexistence have been employed in [107–109]. In [107], coexistence among multiple operators is modeled using conflict graphs, and then a graph coloring technique is applied on the conflict graph. This technique involves dividing a band into multiple orthogonal channels and allocating them to every node in the graph. In [108], multi-channel coloring on a conflict graph based on interference is proposed for a multi-hop cognitive radio network. A distributed graph coloring based channel allocation scheme is discussed in [109], wherein the authors propose a fair channel allocation algorithm for a cognitive radio network.

In literature, *game-theoretic models* are also used to solve the problem of spectrum sharing. In [110] and [111], a game-theoretic approach for solving the coexistence problem has been studied. Two operators dynamically share the spectrum by playing a non-zero sum game in [112]. This work considers both centralized and distributed system models with dynamic spectrum pricing based on demand. In [113], the authors model spectrum sharing among operators as a non-cooperative repeated game. The main concern in the above models is that it may result in an inefficient Nash Equilibrium depending on the operator's utility function. Moreover, the schemes discussed in [112,113] do not guarantee fairness and are also challenging to implement in a realistic scenario.

Most of the above works require sensitive information from the operators while we develop our spectrum sharing algorithm on trivial information. None of the above works discuss the proposed idea of using shared as well as dedicated channel allocation for spectrum sharing. Moreover, we discuss the spectrum sharing problem for a rural setting.

6.2 System Model

We consider an LTE-based middle mile network operating in the TV UHF band. We assume that a portion of this band is available to multiple operators to deploy their networks in rural areas. This portion is divided into multiple orthogonal channels of equal bandwidth. A multi-operator middle mile network architecture is illustrated in Figure 6.1. The network comprises a centralized entity called the 5G-Flow controller that manages LTE-based eNBs[3] in the network. An eNB is the base station in an LTE network and manages mobile users in one or more cells. Multiple low-power eNBs are deployed in a given area, preferably in the vicinity of an optical PoP. Even though an eNB transmits at low power, the coverage area is typically large due to the propagation characteristics of the TV UHF band. An LTE eNB serves multiple CPEs, which in turn provide connectivity to Wi-Fi APs. Here, a CPE act as an LTE based UE. An end-user accesses broadband services through Wi-Fi AP. We consider one CPE per village cluster. Rural areas have clustered settlements as discussed before. Without loss of generality, we assume that these village clusters are uniformly distributed in a given rural area. Consequently, CPEs, with one CPE per cluster, are also uniformly distributed in the given area.

5G-Flow controller along with Spectrum Manager (SM) application allocate channel(s) to various eNBs in the network. The operator registers itself with the SM application to access the TV UHF band. The 5G-Flow controller collects the topology details like antenna height, location, and transmit power of each eNB under an operator and communicates it to the SM. No other details, such as user scheduling information, are shared with the SM for channel allocation. The SM then maintains a database of the information shared by the 5G-Flow controller. The SM treats all operators equally. As we have considered uniform distribution of CPEs in a given area, the average throughput requirement at each eNB is equal. Hence, an operator gives equal priority to all its eNBs. Note that in further discussion, we consider each eNB as an independent network entity. We study the spectrum sharing problem concerning the eNBs, irrespective of the

[3]Although we have considered LTE based network in this chapter, the 5G-Flow controller which is a RAT agnostic control function can manage heterogenous RATs such as 5G NR or Wi-Fi. Therefore this system model can be extented for these RATs.

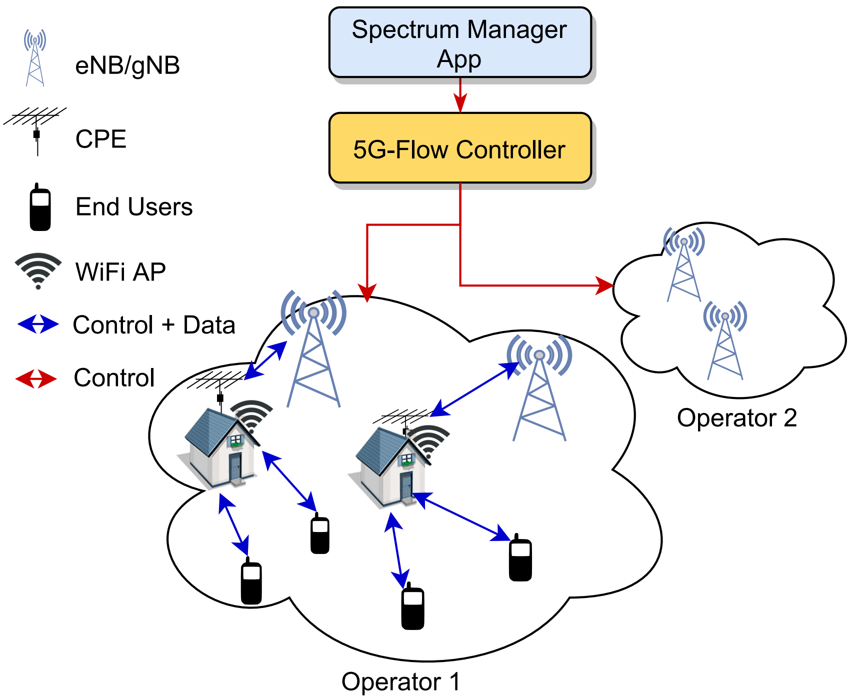


Figure 6.1: Overview of a multi-operator middle mile network

operator. The channel allocated by the SM application is communicated to an eNB of an operator through the 5G-Flow controller.

For simplicity, we consider the *Protocol Interference Model* [114] to model the interference between eNBs. In accordance with this model, two eNBs interfere with each other if they are operating on the same channel and the euclidean distance between them is less than a certain threshold distance. The protocol model formulates an interference state as a binary symmetric matrix, where each element of the matrix indicates whether or not the two eNBs interfere with each other.

6.2.1 Notations

Consider a set $K = \{1, 2, 3, \ldots\}$ representing the eNBs belonging to all the operators in the network. The total number of eNBs is denoted by the cardinality of the set of eNBs, $|K|$. Let $L_k = \{1, 2, 3, \ldots\}$ be the set of $|L_k|$ CPEs served by the eNB$_k, \forall k \in K$. The set of channels available at the SM is given by $M = \{1, 2, 3, \ldots\}$. Also, let $M_k \subset M$ be the set of channels assigned to eNB$_k$. eNB$_k$ allocates resources to its users from the assigned channels.

The SM application allocates channel(s) to the eNBs depending on the interference state of the network. Let $C = \{c_{k,j} | c_{k,j} \in \{0,1\}\}_{|K| \times |K|}$ be a binary symmetric $|K| \times |K|$ matrix where $k, j \in K$, represents the interference state such that:

$$c_{k,j} = \begin{cases} 1, & \text{if eNB}_k \text{ and eNB}_j \text{ interfere with each other,} \\ 0, & \text{otherwise.} \end{cases} \tag{6.1}$$

In addition to allocating channels, SM also defines the mode in which the channel has to be used. The mode of access can be *shared* or *dedicated*. If the mode of a channel assigned to an eNB is dedicated, then that channel does not get allocated to its neighbours. If the mode of access of the assigned channel is shared, then it is to be shared with the neighbours using a sharing mechanism. The channel allocation is given by the two matrices A and B which are defined as follows:

- *Channel Allocation Matrix (A):* $A = \{a_{k,m} | a_{k,m} \in \{0,1\}\}_{|K| \times |M|}$ is a $|K| \times |M|$ binary matrix where, $k \in K$ and $m \in M$ such that:

$$a_{k,m} = \begin{cases} 1, & \text{if channel } m \text{ is assigned to eNB}_k, \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

- *Mode Allocation Matrix (B):* $B = \{b_{k,m} | b_{k,m} \in \{0,1\}\}_{|K| \times |M|}$ is a $|K| \times |M|$ binary matrix where $k \in K$ and $m \in M$. B represents the mode of access on the allocated channel such that:

$$b_{k,m} = \begin{cases} 1, & \text{if allocated channel } a_{k,m} \text{ is to be } \textit{shared}, \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

SM assigns single or multiple channels to an eNB. In the case of multiple channels, they can be contiguous or non-contiguous. When multiple non-contiguous channels are allocated, aggregation is required for cross-channel scheduling. CA, a feature of LTE, is used to serve the purpose of aggregating non-contiguous channels [103]. When the mode of the allocated channel is shared, LBT is used for sharing the channel [104]. LBT is a mechanism in which a radio transmitter performs Clear Channel Assessment (CCA) to check if the medium is idle. The energy in the channel is measured and compared with the detection threshold. If the energy level is greater than the threshold, the channel is assumed to be busy, and the transmitter defers the transmission. If the channel's energy is lower than the threshold, the channel is assumed to be idle, and the transmitter has to back-off for a random number of slots. Here, a slot is a basic unit of time in LBT. Even when the back-off counter reduces to 0, the transmitter can only transmit if the channel is still idle. Once the transmitter gets access to the channel, it can transmit for a fixed duration called *transmit opportunity.* The LBT mechanism has been discussed for the coexistence of LTE and Wi-Fi system [104]. We have used LBT for the coexistence among LTE systems in this work. Note that, since we employ LBT for sharing channels among eNBs, we do not require Inter-Cell Interference Coordination (ICIC) among them.

Once the channel assignment is done by the SM, an eNB allocates resources from the assigned channels to its associated CPEs in a proportional fair manner. The sum throughput at eNB_k is a function of A and B and is given by $T_k(A, B)$. We quantify the fairness F of the system using Jain's fairness index [115] as below:

$$F = \frac{\left(\sum_{k=1}^{|K|} T_k(A, B)\right)^2}{|K| \times \sum_{k=1}^{|K|} T_k(A, B)^2}. \tag{6.4}$$

6.2.2 Problem Formulation

The spectrum sharing problem can be modeled as a system throughput maximization problem under the fairness constraint. Mathematically, the problem can be stated as follows:

$$(A^\star, B^\star) = \underset{A,B}{\arg\max} \left(\sum_{k=1}^{|K|} T_k(A, B)\right), \tag{6.5}$$

$$\text{subject to,} \qquad F \geqslant \delta,$$

$$\text{where,} \qquad \delta = \text{ Constrained value of fairness.}$$

There are two significant challenges in obtaining an optimal solution to this problem. Firstly, this is a combinatorial optimization problem which is known to be NP-complete. Secondly, to determine an optimal solution, a closed-form expression for throughput is required at the eNB. The mathematical expression for LBT throughput can be obtained only for a network that forms a complete graph [116]. In our case, the network graph is not complete. Therefore, in the following section, we propose a heuristic graph-theoretic algorithm to solve the above problem.

6.3 Fairness Constrained Channel Allocation

We employ a graph-theoretic technique to solve the spectrum sharing problem among eNBs. We model the network as a conflict graph $G(V, E)$, where V represents the set of all eNBs, and E denotes the set of edges. An edge between any two eNBs implies that the eNBs interfere with each other, i.e., $E := \{(k, j)|c_{k,j} = 1, \ \forall k, j \in K\}$ where $c_{k,j}$ is an element of the interference matrix, C (defined in Section 6.2.1). In the traditional *Graph Coloring* problem, colors are assigned to the vertices such that vertices with an edge between them do not get the same color. The colors represent the available set of channels denoted by $|M|$. Note that the number of colors (or channels) are considered to be fixed.

We now describe the modifications to the above graph coloring technique for our proposed algorithm. The FCCA algorithm takes graph G as an input. The outputs of the algorithm are the allocation matrices, A and B. Here, G is a graph representing the network as discussed above. In this method, the channels are assigned to the eNBs according to two sub-algorithms described next.

1. *Multiple Dedicated Channel Allocation (MDCA)*: In this sub-algorithm, multiple dedicated channels are assigned to an eNB using greedy graph-coloring method iteratively. It is possible to assign multiple channels to an eNB if the total number of neighbors of an eNB is less than the total number of channels.

2. *One Dedicated Rest Shared Channel Allocation (ODRS-CA)*: In this sub-algorithm, we assign the channels in two steps. In the first step, a single dedicated channel is assigned to each eNB. Then, the set N_k, containing all the channels not assigned to the neighbors of eNB$_k$, is obtained. In the second step, all the channels in N_k are assigned to eNB$_k$ in shared mode.

For a given network topology, the outputs of the sub-algorithms mentioned above are compared to decide the final channel allocation described in Algorithm 4. The algorithm ensures that at least one channel is allocated to each eNB, irrespective of the sub-algorithm being used. Hence, a certain level of fairness is guaranteed by our algorithm. Ideally, the value of δ should be equal to 1 for complete fairness. However, if we give more preference to fairness, the system throughput is compromised. For instance, we observe that constrained values of fairness greater than 0.75 result in a significant decrease in overall network throughput. Hence, we choose the above δ equal to 0.75 to maximize the throughput performance without impairing the fairness.

6.3.1 Example Cases

We now explain the performance of the proposed algorithm using examples. The channel assignment using MDCA algorithm is shown in Figure 6.2 (left) and using ODRS-CA algorithm is shown in Figure 6.2 (right). Bold numbers in the graph denote the dedicated channels; otherwise the allocated channels are to be shared among eNBs. The channel assignment under MDCA algorithm is biased towards vertex 1 as it gets two channels. In contrast, ODRS-CA algorithm gives one dedicated and one shared channel to all. This improves the system fairness with minimal compromise in throughput. The compromise in the throughput occurs due to the sharing of the channel between three vertices.

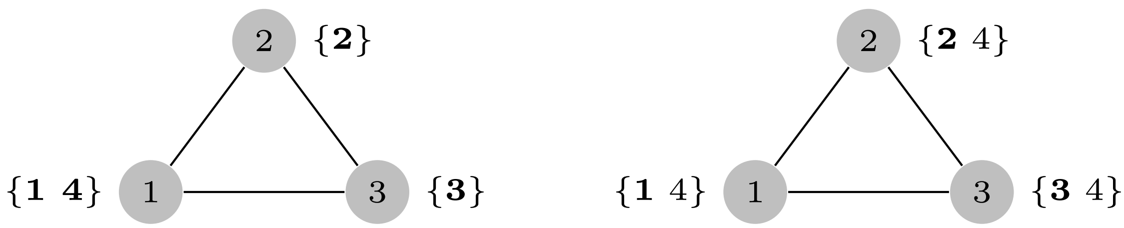


Figure 6.2: Graphs representing the channel allocation for MDCA (left) and ODRS-CA (right) sub-algorithms

We now present a topology where MDCA performs better than ODRS-CA. Consider an example depicted in Figure 6.3. The MDCA algorithm assigns two dedicated channels to all vertices as shown in Figure 6.3 (left), thereby maintaining fairness. The ODRS-CA algorithm is not fair as vertex 2 shares two channels with its neighbors while vertex 1 and 3 get two shared channels which they share with only one neighbor. The overall system throughput is also less in ODRS-CA due to the shared channels. Hence, the MDCA algorithm is more suitable in terms of both throughput and fairness.

Algorithm 4: Fairness Constrained Channel Allocation

Input: Graph G

$\delta = 0.75$

<u>Sub-Algorithm 1 : MDCA</u>

while *N_k is non-empty for all K* **do**

    for *each k from 1 to K* **do**

        Find E_k ; `// set of channels assigned to neighbours of` k

        Obtain $N_k = \{M\} \setminus \{E_k\}$ `// set of feasible channels for eNB`$_k$

        $q \leftarrow \min(N_k)$, $a_{k,q} \leftarrow 1$, $b_{k,q} \leftarrow 0$

    end

end

$T_1 \leftarrow T(A, B)$, $F_1 \leftarrow F(A, B)$, $A_1 \leftarrow A$, $B_1 \leftarrow B$

<u>Sub-Algorithm 2 : ODRS-CA</u>

for *each k from 1 to K* **do**

    Find E_k

    Obtain $N_k = \{M\} \setminus \{E_k\}$

    $q \leftarrow \min N_k$, $a_{k,q} \leftarrow 1$, $b_{k,q} \leftarrow 0$

end

for *each k from 1 to K* **do**

    Find E_k

    Obtain $N_k = \{M\} \setminus \{E_k\}$

    $a_{k,q} \leftarrow 1$, $b_{k,q} \leftarrow 1 \quad \forall q \in N_k$

end

$T_2 \leftarrow T(A, B)$, $F_2 \leftarrow F(A, B)$, $A_2 \leftarrow A$, $B_2 \leftarrow B$

Check F_1 and F_2 and choose $(A^\star, B^\star)$ such that the fairness is greater than δ. If both are greater than δ then choose $(A^\star, B^\star)$ corresponding to $\max(T_1, T_2)$.

Output: return A^*, B^*

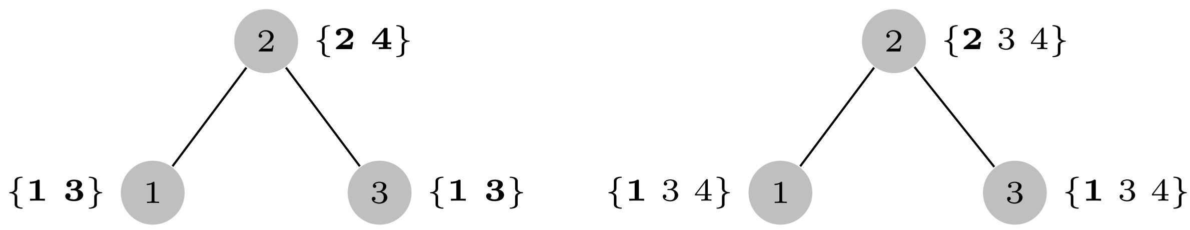


Figure 6.3: Graphs representing the channel allocation for MDCA (left) and ODRS-CA (right) sub-algorithms

As seen by the above two contrasting examples, one of the two sub-algorithms performs better depending on the system topology. Hence, we use a combination of two sub-algorithms to enhance system performance.

6.4 Performance Analysis

In this section, we present the results of system-level simulations (in ns-3) to assess the performance of FCCA algorithm. We also compare the proposed approach with few other coexistence approaches.

6.4.1 Coverage Radius of eNB operating in TV UHF band

The coverage radius is defined as the maximum allowed distance between the transmitter and the receiver to enable communication. We calculate the coverage radius of an eNB using the following link budget equation:

$$RS = P_t + G_t + G_r - PL(d_0, h_t, h_r, f_c) - C - NF, \tag{6.6}$$

where RS is receiver sensitivity as specified for an LTE-A system. P_t, G_t and G_r are the transmit power (in dBm), the transmitter antenna gain (in dBi) and the receiver antenna gain (in dBi), respectively. $PL(d, h_t, h_r, f_c)$ is the path loss which is a function of distance d between transmitter and receiver, antenna heights (h_t and h_r) and carrier frequency (f_c). Hata model for Suburban Areas is used to calculate path loss [117]. The transmit power, P_t, of an eNB is kept low at 18 dBm. C is the cable loss and NF is the receiver noise figure. The above equation can be used to calculate the maximum allowed distance d_0 between eNB and CPE depending on the minimum receiver sensitivity (RS) as specified in LTE standard. For the values of the parameters given in Table 6.1, the coverage radius of eNB is approximately 3 km.

Table 6.1: Simulation Parameters

Parameters	**Values**
Central Frequency(f_c)	500-520 MHz
Transmit Power (P_t)	18 dBm
Receiver Sensitivity (RS)	-101 dBm [118]
Cable Loss (C)	2 dB
Receiver Noise Figure (NF)	7 dB
Transmitter Antenna Gain (G_t)	10 dB
Receiver Antenna Gain (G_r)	0 dB
Transmitter Antenna Height (h_t)	30 m
Receiver Antenna Height (h_r)	5 m
Slot Time	9 μs
Transmit Opportunity	10 ms
Detection Threshold	-62 dBm
Simulation time	30 s

6.4.2 Simulation Scenario

We assume that 20 MHz of the TV UHF band is available for the middle mile network. This band is further divided into four orthogonal channels of 5 MHz each. Without loss of generality, we assume all channels to be identical. As the rural areas are sparsely populated, we assume that an eNB will not experience interference from more than three neighboring eNBs. The eNBs are deployed uniformly at random in an area of 100 sq. km as shown in Figure 6.4. Each eNB has a coverage radius of 3 km as calculated above. The CPEs are distributed uniformly in the given area. Each eNB is assumed to serve 5 stationary CPEs. For constructing the conflict graph using the protocol interference model, we consider the threshold as 4 km distance between two eNBs. If the distance between eNBs is less than 4 km, then they interfere with each other. We perform ns-3 simulations over 100 random topologies. All the performance metrics are averaged over such realizations. The simulation parameters are given in Table 6.1. We do not consider fast fading effects in our simulation scenario. We consider only saturated downlink transmission, i.e., at each eNB, saturated traffic is generated for each of the associated CPEs.

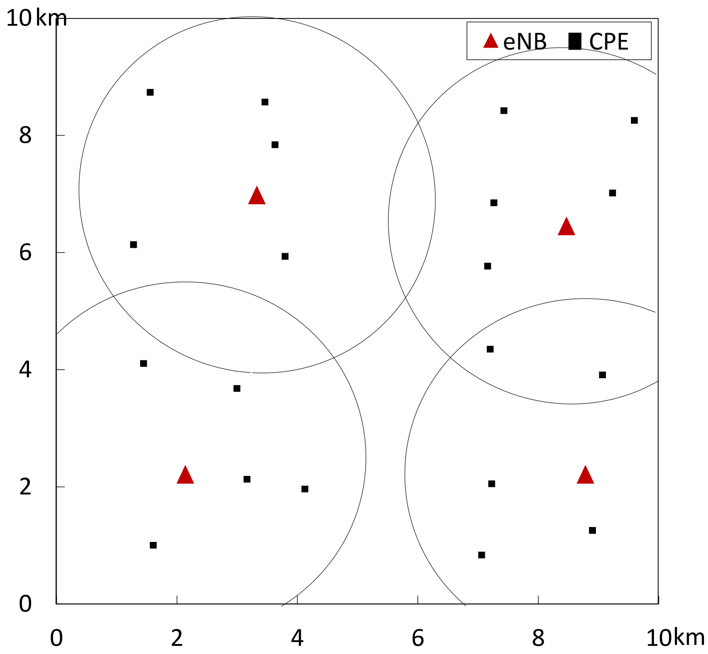
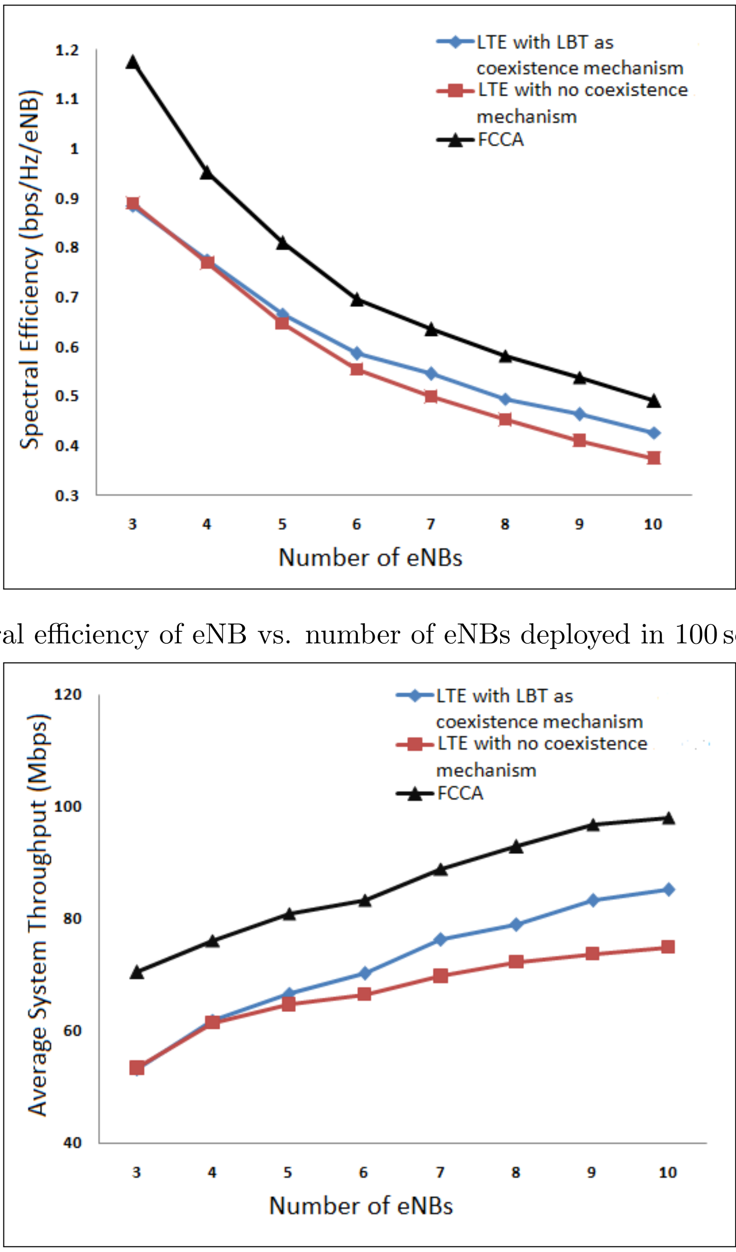


Figure 6.4: An example topology of the network

6.4.3 Results

We analyze three performance metrics to assess the performance of the proposed FCCA algorithm — i) *Spectral efficiency*, ii) *Average system throughput per eNB*, and iii) *Jain's fairness index*. The performance of these metrics is observed with respect to an increase in the network density, wherein we increase the number of eNBs from 3 to 10 in a fixed area of 100 sq. km. In our results, we measure spectral efficiency per eNB in bits/s/Hz. We have used Jain's fairness index to quantify how fairly the available band is shared among the eNBs.

In Figure 6.5, we compare the spectral efficiency and the average system throughput of FCCA with two other schemes — i) LTE with no-coexistence mechanism, and ii) LTE with LBT as the coexistence mechanism. In the first scheme, the entire 20 MHz band is used by all eNBs without any coexistence mechanism. As illustrated in the graph, the spectral efficiency is low due to interference among the eNBs. With LBT as the coexistence mechanism, the entire 20 MHz band is shared among all the eNBs. As a result, the performance is affected as the transmission time is wasted in contention. The FCCA algorithm performs better than the above schemes as it considers the network topology for allocating the channels. As shown in Figure 6.6, the fairness under the FCCA



(a) Spectral efficiency of eNB vs. number of eNBs deployed in 100 sq. km area

(b) Average throughput per eNB vs. number of eNBs deployed in 100 sq. km area

Figure 6.5: Comparative analysis of average system throughput and spectral efficiency i) LTE with no coexistence mechanism ii) LTE with LBT as coexistence mechanism and iii) FCCA

algorithm is also better than other two schemes. The proposed algorithm guarantees a fairness index of 0.76 even when there are 10 eNBs per 100 sq. km.

It is crucial to analyze the performance of the proposed algorithm under the best and the worst case scenario. Consider a very sparse network (best case) where only three

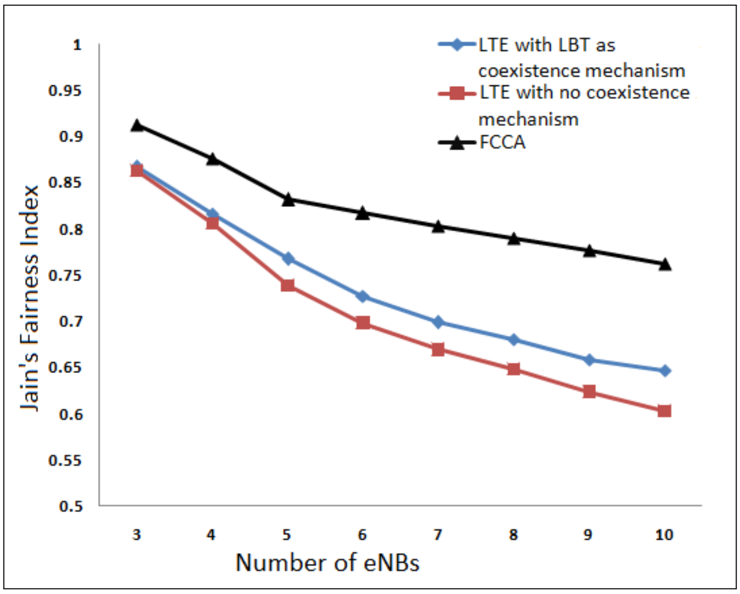


Figure 6.6: Comparative analysis of Jain's fairness index for i) LTE with no coexistence mechanism ii) LTE with LBT as coexistence mechanism and iii) FCCA

eNBs are deployed in the given area. When 100 random topologies are simulated under this setting, it is observed that MDCA and ODRS-CA are equally preferred. This result highlights the fact that orthogonal channel allocation may not always give the best result. A combination of orthogonal and shared channel allocation improves system performance. In a very dense network (worst case), wherein 10 eNBs are deployed in the given area, ODRS-CA is preferred between the two sub-algorithms in 61% cases. This result further underscores the need to use a combination of shared and dedicated channel allocation as a coexistence mechanism for better system performance in terms of spectral efficiency and fairness.

6.5 Summary

Multi-operator resource sharing is an important characteristic of Frugal 5G network. Spectrum, a highly expensive resource, if shared among multiple operators, the network costs can be reduced significantly. In this chapter, we have presented a graph-theoretic channel allocation algorithm to enable the spectrum sharing among multiple middle mile networks owned by same or different operators. We have proposed a novel concept of allocating a combination of shared and dedicated channel to an eNB. The performance of the algorithm has been studied using ns-3 simulations. We show that when all the available channels are allocated in dedicated manner or when all the channels are shared, the performance is affected. When a combination of shared and dedicated channels is used, as in our proposed algorithm, metrics such as spectral efficiency and fairness among operators both improve significantly.

Chapter 7

Conclusion and Future Work

Although the Internet has transformed many spheres of our everyday lives, a large section of the global population is unconnected and hence unable to experience these changes. A majority of unconnected individuals belong to the rural areas of developing countries and LDCs, resulting in rural-urban digital divide. To address this digital divide, we have analyzed the existing connectivity technologies and research literature. Accordingly, we have observed that these proposals suggest a piecemeal solution to the rural broadband problem. To enable a comprehensive solution, we have articulated the challenges and requirements relating to rural broadband based on the learnings from the deployment of rural broadband testbeds as part of a project at IIT Bombay. We have also highlighted the need to develop a novel network architecture for rural Internet access. In this thesis, we propose a novel network architecture with detailed implementation perspectives and performance comparisons. We make the following conclusions based on the proposals presented in this thesis.

- **Frugal 5G network architecture for rural areas**: Frugal 5G is one of the first network architectures to provide a comprehensive solution to the rural broadband problem, taking into account various characteristics and requirements pertaining to connectivity in rural areas. Frugal 5G is a heterogeneous network architecture along with a fog element that control and manages the network. The proposed network architecture addresses rural connectivity concerns through its novel features—unified control for a multi-RAT network at the edge, unified interworking function, remote operation and localized communication, flexible fog deployment, and framework for network virtualization and network slice creation. Qualitative analysis has also demonstrated that the proposed network architecture successfully meets the requirements of rural areas.

- **5G-Flow: Realization of Frugal 5G network architecture**: To take the Frugal 5G network architecture towards adoption, we have implemented the network architecture using existing technologies such as 3GPP 5G and Wi-Fi networks. 5G-Flow network supplements 3GPP 5G network with OpenFlow, successfully realizing a network architecture that features unified multi-RAT RAN, unified interworking function and flexible fog deployment. We do not discuss multi-operator resource sharing in detail as that can natively be supported by the underlying 5G network. We have analyzed the impact of the proposed enhancements on the existing 3GPP 5G network, suggesting a minimal impact involving software-based changes only. We have also demonstrated additional use-cases of 5G-Flow network.

- **Quantitaive analysis of 5G-Flow network**: To analyze the performance of the 5G-Flow network quantitatively, we have developed a 5G multi-RAT system-level simulator in MATLAB and released the source code under MIT license for the benefit of the research community. We have built a flexible simulator to enable the implementation of novel multi-RAT network architectures. The results show that the 5G-Flow network uniformly balances the load across multiple RATs in the downlink scenario supporting multiple service types. Due to the uplink/downlink decoupling, the overall network performance and, in particular, the Wi-Fi network performance have also improved.

- **Middle mile network planning**: Efficient planning of a large-scale network such as the middle mile network can indeed impact infrastructure utilization and system performance. The results of our proposed simulated annealing based heuristic algorithm emphasize the stated concept by demonstrating that it minimizes the network infrastructure cost by optimizing the number of nodes and hops in the middle mile network.

- **Spectrum sharing for multi-operator middle mile network**: Resource sharing is an essential aspect of the Frugal 5G network, and we explore this aspect for multiple middle mile networks owned by same or different operators. Our proposed FCCA algorithm highlights the need to use a combination of shared and dedicated channel allocation to eNBs for improved spectral efficiency and fairness among operators in a network.

7.1 Bringing it All Together: IEEE P2061 Standardization

A standard defines "requirements, specifications, guidelines or characteristics for a determined material, product, process or service" [119]. A successful standard can influence regulations as well as accelerate the industrial adoption of the technology. Standardization can also reduce risks, costs, and development time for a technology. Therefore, an impactful solution towards any problem must focus on standardization in addition to innovation. We have taken a step in this direction by presenting the proposed solutions towards rural broadband problem to an ongoing standards development project.

IEEE has initiated a working group to standardize a reference architecture for rural areas. Under this working group, a standards development project, P2061, aims to propose "Architecture for Low Mobility Energy Efficient Network for Affordable Broadband Access" [46]. P2061 working group has members from diverse backgrounds, including infrastructure providers, equipment manufacturers, and operators. Moreover, the members are scattered around the world, resulting in exchange of different perspectives regarding the problem.

P2061 standard is currently being developed, and we have been actively participating in the standardization work. We have made three key contributions in the working group meetings — Rural requirements/characteristics, Frugal 5G Network Architecture, and 5G-Flow Network. These contributions have been discussed in detail by the working group. As these contributions suggest a network architecture for rural areas that can be implemented using 3GPP 5G network, industrial adoption seems probable. These contributions may form a significant part of the final draft of the P2061 standard.

7.2 Future Work

Enabling connectivity in rural areas is an enormous task and we have addressed some major aspects of this problem in the thesis. However, following are some open research issues that need to be examined.

7.2.1 Economic Analysis of the Frugal 5G Network

The Frugal 5G network presents various essential features to meet the rural connectivity requirements. One of the critical requirements of any rural area is affordability. In this thesis, we have provided a qualitative analysis explaining the various features which help in reducing the network costs. We have also provided a rough estimate for the infrastructure cost for the Frugal 5G network. However, it is important to conduct a comprehensive cost analysis to demonstrate the affordability and economic viability of the proposed network architecture.

7.2.2 Algorithms and Prototype Development for 5G-Flow network

5G-Flow network utilizes the 3GPP 5G network framework to realize important network design characteristics that enable an efficient rural broadband network. One of the desired characteristics is providing *localized communication* to people in rural areas even when the backhaul is absent, which needs to be addressed in the 5G-Flow network. The research issue here is how to provide the user-level information at the access network so that appropriate traffic routing can be done at the edge. Also, developing efficient routing algorithms for local routing of data need to be developed. In future work, we aim to address these issues.

To make the 5G-Flow network available for use in rural areas, it is also important to develop its working prototype. We plan to develop and test the prototype on the 5G testbed developed in our lab at IIT Bombay. The prototype will help us in analyzing and further improving the 5G-Flow network.

7.2.3 Flexible 5G simulator

A system-level simulator is a vital tool to analyze the behavior of any communication network. Additionally, one can evaluate new features and plan the network using system-level simulators. Therefore, it is required to design a system-level simulator that allows for easier implementation of new network architectures for 5G and beyond. We aim to develop the 5G multi-RAT simulator further, enabling the testing of software-defined networking and virtualization based concepts. Moreover, the simulator will provide appropriate network abstractions for the said objective.

7.2.4 Potential Network Architecture for 6G

ITU has established a Focus Group for Network 2030 (FG NET-2030) to study the capabilities of the network to support emerging applications and business cases [120]. One of the key focus areas of Network 2030 is to enable ubiquitous connectivity. In addition to this, Network 2030 aims to support high-precision networks such as holographic-type communications. FG NET-2030 discusses the gaps in the existing networks and potential solutions to enable these applications in the 6G network.

For enabling these emerging applications, a shift is required from a centralized architecture (of 5G) to a flat and distributed network architecture, supporting ultra-low latency and immense scalability. Decoupling of access network from the core network and making connectivity to core network optional can be a significant step towards 6G. Moreover, features like multi-RAT integration at access network may also be crucial for future networks.

Although the network architectures in this thesis have been proposed in context of rural areas, the proposals also address some of the above-stated requirements of 6G network and can be further enhanced to suggest an alternative for the 6G network architecture.

Bibliography

[1] J. Culkin, "A schoolman's guide to Marshall McLuhan," *Saturday Review*, pp. 51–53, 1967.

[2] M. McLuhan and B. R. Powers, "The global village: Transformations in world life and media in the 21st century," *Oxford University Press*, 1989.

[3] T. Freidman, "The world is flat," *New York: Farrar, Straus and Giroux*, vol. 488, 2005.

[4] WSIS-03/GENEVA/DOC/4-E, "Declaration of Principles," *World Summit on the Information Society*, December 2003.

[5] "ICT Facts and Figures," International Telecommunication Union, Report, 2019. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf

[6] "Indian Telecom Services Performance Indicator Report," Telecom Regulatory Authority of India, Report, 2020. [Online]. Available: https://trai.gov.in/sites/default/files/PIR_17092020_0.pdf

[7] K. Sprague, F. Grijpink, J. Manyika, L. Moodley, B. Chappuis, K. Pattabiraman, and J. Bughin, "Offline and falling behind: Barriers to internet adoption," McKinsey & Company, Technical Report, 2014.

[8] "Extension, Department of Agriculture, Cooperation and Farmers Welfare," Website, [Date accessed: 04.03.2021]. [Online]. Available: http://agricoop.gov.in/divisiontype/extension

[9] "Level up: Need for deeper fibre penetration," Tele.net, January 2020, [Date accessed: 04.03.2021]. [Online]. Available: https://tele.net.in/level-up-need-for-deeper-fibre-penetration/

[10] "Bharat Broadband Network (BharatNet)," Website, [Date accessed: 04.03.2021]. [Online]. Available: www.bbnl.nic.in

[11] IEEE Std 802.11-2016, "IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," pp. 1–3534, December 2016.

[12] B. Raman and K. Chebrolu, "Experiences in using WiFi for rural internet in India," *IEEE Communications Magazine*, vol. 45, pp. 104–110, January 2007.

[13] R. Patra, S. Nedevschi, S. Surana, A. Sheth, L. Subramanian, and E. Brewer, "Wildnet: Design and implementation of high performance wifi based long distance networks," in *USENIX Conference on Networked Systems Design*, April 2007, pp. 87–100.

[14] F. Darbari, M. Brew, S. Weiss, and W. S. Robert, "Practical aspects of broadband access for rural communities using a cost and power efficient multi-hop/relay network," in *IEEE Global Communications Conference (GLOBECOM) Workshops*, December 2010, pp. 731–735.

[15] J. Backens, G. Mweemba, and G. Van Stam, "A Rural Implementation of a 52 Node Mixed Wireless Mesh Network in Macha, Zambia," in *International Conference on e-Infrastructure and e-Services for Developing Countries*, December 2010, pp. 32–39.

[16] B. Raman and K. Chebrolu, "Design and Evaluation of a New MAC Protocol for Long-distance 802.11 Mesh Networks," in *International Conference on Mobile Computing and Networking (MobiCom)*, August 2005, pp. 156–169.

[17] L. Subramanian, S. Nedevschi, M. Ho, E. Brewer, and A. Sheth, "Rethinking Wireless for the Developing World," in *Fifth Workshop on Hot Topics in Networks (HotNets-V)*, November 2006.

[18] S. Sen and B. Raman, "Long Distance Wireless Mesh Network Planning: Problem Formulation and Solution," in *International Conference on World Wide Web*, May 2007, pp. 893–902.

[19] P. Dutta, S. Jaiswal, D. Panigrahi, and R. Rastogi, "A New Channel Assignment Mechanism for Rural Wireless Mesh Networks," in *IEEE Conference on Computer Communications (INFOCOM)*, April 2008, pp. 2261–2269.

[20] "Consultation Paper on Spectrum related issues: efficient utilization, spectrum allocation, and spectrum pricing," Telecom Regulatory Authority of India, Report. [Online]. Available: https://www.trai.gov.in/sites/default/files/201506090333529087238siddharth.pdf

[21] G. Naik, S. Singhal, A. Kumar, and A. Karandikar, "Quantitative assessment of TV white space in India," in *Twentieth National Conference on Communications (NCC)*, February 2014, pp. 1–6.

[22] C. S. Sum, G. P. Villardi, M. A. Rahman, T. Baykas, H. N. Tran, Z. Lan, C. Sun, Y. Alemseged, J. Wang, C. Song, C. W. Pyo, S. Filin, and H. Harada, "Cognitive communication in TV white spaces: An overview of regulations, standards, and technology," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 138–145, July 2013.

[23] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, January 2009.

[24] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, "IEEE 802.11af: A standard for TV white space spectrum sharing," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92–100, October 2013.

[25] K. Kazaura, K. Wakamori, M. Matsumoto, T. Higashino, K. Tsukamoto, and S. Komaki, "RoFSO: a universal platform for convergence of fiber and free-space optical communication networks," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 130–137, January 2010.

[26] K. W. Matthee, G. Mweemba, A. V. Pais, G. van Stam, and M. Rijken, "Bringing Internet connectivity to rural Zambia using a collaborative approach," in *International Conference on Information and Communication Technologies and Development*, December 2007, pp. 1–12.

[27] "Terragraph–virtual fiber for high speed fixed broadband," Telecom Infra Project, White Paper, 2018, [Date accessed: 04.03.2021]. [Online]. Available: https://terragraph.com/wp-content/uploads/2019/02/Terragraph_MikebudaTrial_Whitepaper.pdf

[28] IEEE Std 802.16.1-2012, "IEEE Standard for WirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems," pp. 1–1090, September 2012.

[29] StatCounter, Web Traffic Analysis Website. [Date accessed: 04.03.2021]. [Online]. Available: https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/

[30] "Jio 3G / 4G / 5G coverage map, India," nPerf, Website, [Date accessed: 04.03.2021]. [Online]. Available: https://www.nperf.com/en/map/IN/-/19111.Jio/signal/

[31] M. Series, "IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond," Recommendation ITU 2083, Technical Specification, 2015.

[32] 3GPP TS 23.501 v16.7.0, "System Architecture for the 5G System," Technical Specification, 2020. [Online]. Available: https://www.3gpp.org/DynaReport/23501.htm

[33] A. Karlsson, O. Al-Saadeh, A. Gusarov, R. V. R. Challa, S. Tombaz, and K. W. Sung, "Energy-efficient 5G deployment in rural areas," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, December 2016, pp. 1–7.

[34] L. Chiaraviglio, N. Blefari-Melazzi, W. Liu, J. A. Gutierrez, J. van de Beek, R. Birke, L. Chen, F. Idzikowski, D. Kilper, P. Monti, A. Bagula, and J. Wu, "Bringing 5g

into rural and low-income areas: Is it feasible?" *IEEE Communications Standards Magazine*, vol. 1, no. 3, pp. 50–57, October 2017.

[35] M. Khalil, J. Qadir, O. Onireti, M. A. Imran, and S. Younis, "Feasibility, architecture and cost considerations of using TVWS for rural Internet access in 5G," in *IEEE Conference on Innovations in Clouds, Internet and Networks (ICIN)*, March 2017, pp. 23–30.

[36] A. Kumar, A. Karandikar, G. Naik, M. Khaturia, S. Saha, M. Arora, and J. Singh, "Toward enabling broadband for a billion plus population with TV white spaces," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 28–34, July 2016.

[37] M. Khaturia, S. B. Belur, and A. Karandikar, "Chapter 5 - TV White Space Technology for Affordable Internet Connectivity," in *TV White Space Communications and Networks*. Woodhead Publishing, 2018, pp. 83–96.

[38] M. Khaturia, J. Singh, A. Patil, S. B. Belur, A. Karandikar, P. Chaporkar, and T. Ghadge, "Towards Frugal 5G: A Case Study of Palghar Test-bed in India," in *IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, March 2019, pp. 1–6.

[39] "All India Rural Financial Inclusion Survey 2016-17," NABARD, Survey Report, [Date accessed: 04.03.2021]. [Online]. Available: https://www.nabard.org/auth/writereaddata/tender/1608180417NABARD-Repo-16_Web_P.pdf

[40] *Rural Land Area of India*, The World Bank. [Online]. Available: https://data.worldbank.org/indicator/AG.LND.TOTL.RU.K2?end=2010&locations=IN&start=2010

[41] *Local Government Directory*, Ministry of Panchayati Raj, Government of India. [Online]. Available: https://lgdirectory.gov.in/

[42] A. Marcus and A. Wong, "Internet for All: A Framework for Accelerating Internet Access and Adoption," in *World Economic Forum*, May 2016, pp. 4–5.

[43] P. Fan, J. Zhao, and C. I, "5G high mobility wireless communications: Challenges and solutions," *China Communications*, vol. 13, pp. 1–13, February 2016.

[44] "Progress report of village electrification," Ministry of Power, Government of India, Central Electricity Authority Report, 2017, [Date accessed 25.05.2019]. [Online]. Available: https://data.gov.in/catalog/progress-report-village-electrification

[45] 5G-Flow Simulator Source Code. [Online]. Available: https://github.com/MeghnaKhaturia/5G-Flow-RAN

[46] P2061, "Architecture for Low Mobility Energy Efficient Network for Affordable Broadband Access," IEEE Standards Development Project, 2018. [Online]. Available: https://standards.ieee.org/project/2061.html

[47] ITU-R M.2411, "Requirements, evaluation criteria and submission templates for the development of IMT-2020," International Telecommunications Union, Technical Report, 2017. [Online]. Available: https://www.itu.int/pub/R-REP-M.2411

[48] ITU-R M.2412, "Guidelines for evaluation of radio interface technologies for IMT-2020," International Telecommunications Union, Technical Report, 2017. [Online]. Available: https://www.itu.int/pub/R-REP-M.2412

[49] ETSI GSNFV v1.2.1, "Network functions virtualisation (NFV): Architectural framework," Group Specification, 2014. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01.02.01_60/gs_NFV002v010201p.pdf

[50] OpenFog Consortium Architecture Working Group, "OpenFog Reference Architecture for Fog Computing," Technical Specification, 2017. [Online]. Available: https://www.openfogconsortium. org/wp-content/uploads/OpenFog_Reference_Architecture_2 _09_17-FINAL.pdf

[51] S. Denazis, E. Haleplidis, J. H. Salim, O. Koufopavlou, D. Meyer, and K. Pentikousis, "RFC 7426 - Software-Defined Networking (SDN): Layers and Architecture Terminology," Technical Specification, 2015.

[52] 3GPP TS 38.473 v16.4.0, "F1 Application Protocol (F1-AP)," Technical Specification, 2021. [Online]. Available: https://www.3gpp.org/DynaReport/38473.htm

[53] K. Johansson, A. Furuskar, P. Karlsson, and J. Zander, "Relation between base station characteristics and cost structure in cellular systems," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 4, 2004, pp. 2627–2631.

[54] *UniFi AP Outdoor*. [Online]. Available: https://www.ui.com/unifi/unifi-ap-outdoor/

[55] *PowerBeam AC Gen2*. [Online]. Available: https://www.ui.com/airmax/powerbeam-ac-gen2/

[56] "Recommendations on National Broadband Plan," Telecom Regulatory Authority of India, Report. [Online]. Available: https://www.trai.gov.in/sites/default/files/Rcommendation81210.pdf

[57] T. De Schepper, P. Bosch, E. Zeljković, F. Mahfoudhi, J. Haxhibeqiri, J. Hoebeke, J. Famaey, and S. Latré, "ORCHESTRA: Enabling Inter-Technology Network Management in Heterogeneous Wireless Networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 4, pp. 1733–1746, December 2018.

[58] K. K. Yap, M. Kobayashi, R. Sherwood, T. Y. Huang, M. Chan, N. Handigol, and N. McKeown, "OpenRoads: Empowering Research in Mobile Networks," *SIGCOMM Computer Communication Review (CCR)*, vol. 40, no. 1, pp. 125–126, January 2010.

[59] K. K. Yap, M. Kobayashi, D. Underhill, S. Seetharaman, P. Kazemian, and N. McKeown, "The stanford openroads deployment," in *ACM International Workshop on Experimental Evaluation and Characterization*, September 2009, pp. 59–66.

[60] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng, "OpenRAN: a software-defined ran architecture via virtualization," *ACM SIGCOMM Computer Communication Review (CCR)*, vol. 43, no. 4, pp. 549–550, August 2013.

[61] A. N. Manjeshwar, A. Roy, P. Jha, and A. Karandikar, "Control and Management of Multiple RATs in Wireless Networks: An SDN Approach," in *IEEE 2nd 5G World Forum (5GWF)*, October 2019, pp. 596–601.

[62] E. Coronado, S. N. Khan, and R. Riggio, "5G-EmPOWER: A Software-Defined Networking Platform for 5G Radio Access Networks," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 715–728, June 2019.

[63] P. K. Taksande, P. Jha, A. Karandikar, and P. Chaporkar, "Open5G: A Software-Defined Networking Protocol for 5G Multi-RAT Wireless Networks," in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, April 2020, pp. 1–6.

[64] S. Chandrashekar, A. Maeder, C. Sartori, T. Höhne, B. Vejlgaard, and D. Chandramouli, "5G multi-RAT multi-connectivity architecture," in *IEEE International Conference on Communications Workshops (ICC)*, May 2016, pp. 180–186.

[65] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: Software defined radio access network," in *Second ACM SIGCOMM workshop on Hot topics in Software Defined Networking (HotSDN)*, August 2013, pp. 25–30.

[66] T. Chen, H. Zhang, X. Chen, and O. Tirkkonen, "SoftMobile: Control evolution for future heterogeneous mobile networks," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 70–78, December 2014.

[67] X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and K. Kontovasilis, "FlexRAN: A flexible and programmable platform for software-defined radio access networks," in *Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, December 2016, pp. 427–441.

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

[69] C. J. Bernardos, A. De La Oliva, P. Serrano, A. Banchs, L. M. Contreras, H. Jin, and J. C. Zúñiga, "An architecture for software defined wireless networking," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 52–61, June 2014.

[70] I. F. Akyildiz, P. Wang, and S.-C. Lin, "SoftAir: A software defined networking architecture for 5G wireless systems," *Computer Networks*, vol. 85, pp. 1–18, January 2015.

[71] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "SoftCell: Scalable and Flexible Cellular Core Network Architecture," in *Ninth ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, December 2013, pp. 163–174.

[72] M. Bansal, J. Mehlman, S. Katti, and P. Levis, "Openradio: a programmable wireless dataplane," in *First workshop on Hot topics in Software Defined Networks (HotSDN)*, August 2012, pp. 109–114.

[73] K. Pentikousis, Y. Wang, and W. Hu, "Mobileflow: Toward software-defined mobile networks," *IEEE Communications magazine*, vol. 51, no. 7, pp. 44–53, 2013.

[74] L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, and T. Vazao, "Towards programmable enterprise WLANs with Odin," in *First workshop on Hot topics in Software Defined Networks (HotSDN)*, August 2012, pp. 115–120.

[75] D. Laselva, D. Lopez-Perez, M. Rinne, and T. Henttonen, "3GPP LTE-WLAN Aggregation Technologies: Functionalities and Performance Comparison," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 195–203, March 2018.

[76] P. Nuggehalli, "LTE-WLAN aggregation [Industry Perspectives]," *IEEE Wireless Communications*, vol. 23, no. 4, pp. 4–6, August 2016.

[77] "OpenFlow Switch Specification, v1.3.0," Open Networking Foundation, Technical Specification, 2014. [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-spec-v1.3.0.pdf

[78] R. Enns, M. Bjorklund, J. Schoenwaelder, and A. Bierman, "RFC 6241 - Network configuration protocol (NETCONF)," Technical Specification, June 2011. [Online]. Available: https://tools.ietf.org/html/rfc6241

[79] "OpenFlow Management and Configuration Protocol, v1.2," Open Networking Foundation, Technical Specification, 2014. [Online]. Avail-

able: https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow-config/of-config-1.2.pdf

[80] "OpenFlow Notifications Framework, OpenFlow Management, v1.0," Open Networking Foundation, Technical Specification, 2013. [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2013/02/of-notifications-framework-1.0.pdf

[81] 3GPP TS 24.193 v16.2.0, "5G System; Access Traffic Steering, Switching and Splitting (ATSSS)," Technical Specification, 2020. [Online]. Available: https://www.3gpp.org/DynaReport/24193.htm

[82] 3GPP TS 38.211 v16.4.0, "NR; Physical channels and modulation," Technical Specification, 2021. [Online]. Available: https://www.3gpp.org/DynaReport/38211.htm

[83] W. Stallings, "Operating systems: internals and design principles (7th ed.)," 2012.

[84] 3GPP TR 38.901 v16.1.0, "Study on channel model for frequencies from 0.5 to 100 GHz," Technical Report, 2020. [Online]. Available: https://www.3gpp.org/DynaReport/38.901.htm

[85] 3GPP TS 38.214 v16.4.0, "NR; Physical layer procedures for data," Technical Specification, 2021. [Online]. Available: https://www.3gpp.org/DynaReport/38214.htm

[86] ITU-R M. 2135, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," Internation Telecommunications Union, Technical Specification, 2008. [Online]. Available: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2135-1-2009-PDF-E.pdf

[87] "Cisco Aironet 1570 outdoor Wi-Fi AP," Datasheet. [Online]. Available: https://www.cisco.com/c/en/us/products/wireless/aironet-1570-series/index.html

[88] J. Terry and J. Heiskala, "OFDM wireless LANs: A theoretical and practical guide (2nd ed.)," *Sams publishing*, 2002.

[89] 3GPP TS 38.331 v16.3.1, "5G; NR; Radio Resource Control (RRC); Protocol specification," Technical Specification, 2021. [Online]. Available: https://www.3gpp.org/dynareport/38331.htm

[90] Y. T. Hou, Y. Shi, H. D. Sherali, and S. F. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2579–2590, November 2005.

[91] E. L. Lloyd and G. Xue, "Relay node placement in wireless sensor networks," *IEEE Transactions on Computers*, vol. 56, no. 1, pp. 134–138, June 2007.

[92] M. N. Islam, A. Sampath, A. Maharshi, O. Koymen, and N. B. Mandayam, "Wireless backhaul node placement for small cell networks," in *Information Sciences and Systems (CISS)*, May 2014, pp. 1–6.

[93] K. Nitesh and P. K. Jana, "Relay node placement algorithm in wireless sensor network," in *IEEE International Advance Computing Conference (IACC)*, March 2014, pp. 220–225.

[94] Y. Yu, S. Murphy, and L. Murphy, "Planning base station and relay station locations for IEEE 802.16j network with capacity constraints," in *IEEE Consumer Communications and Networking Conference (CCNC)*, February 2010, pp. 1–5.

[95] L. C. Wang, W. S. Su, J. H. Huang, A. Chen, and C. J. Chang, "Optimal relay location in multi-hop cellular systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2008, pp. 1306–1310.

[96] Z. Abichar, A. E. Kamal, and J. M. Chang, "Planning of relay station locations in IEEE 802.16 (WiMAX) networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2010, pp. 1–6.

[97] H. C. Lu and W. Liao, "Joint base station and relay station placement for IEEE 802.16j networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, March 2009, pp. 1–5.

[98] H. C. Lu, W. Liao, and F. Y. S. Lin, “Relay station placement strategy in IEEE 802.16j WiMAX networks,” *IEEE Transactions on Communications*, vol. 59, no. 1, pp. 151–158, November 2011.

[99] C. Y. Chang, C. T. Chang, M. H. Li, and C. H. Chang, “A novel relay placement mechanism for capacity enhancement in IEEE 802.16j WiMAX networks,” in *IEEE International Conference on Communications (ICC)*, August 2009, pp. 1–5.

[100] C. Y. Chang and M. H. Li, “A placement mechanism for relay stations in 802.16j WiMAX networks,” *Wireless networks*, vol. 20, no. 2, pp. 227–243, May 2014.

[101] D. Yang, X. Fang, G. Xue, and J. Tang, “Relay Station Placement for Cooperative Communications in WiMAX Networks,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–5.

[102] “MATLAB 2020a.” [Online]. Available: www.mathworks.com

[103] 3GPP TR 36.808 v10.1.0, “Evolved Universal Terrestrial Radio Access (E-UTRA); Carrier Aggregation; Base Station (BS) radio transmission and reception,” Technical Specification, 2013. [Online]. Available: https://www.3gpp.org/DynaReport/36808.htm

[104] 3GPP TS 36.889 v13.0.0, “Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum,” Technical Specification, 2015. [Online]. Available: https://www.3gpp.org/DynaReport/36889.htm

[105] A. Alsohaily and E. S. Sousa, “Spectrum sharing LTE-advanced small cell systems,” in *International Symposium on Wireless Personal Multimedia Communications (WPMC)*, June 2013, pp. 1–5.

[106] Y. Teng, Y. Wang, and K. Horneman, “Co-primary spectrum sharing for denser networks in local area,” in *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, June 2014, pp. 120–124.

[107] S. Sengupta, S. Brahma, M. Chatterjee, and N. S. Shankar, “Self-coexistence among interference-aware IEEE 802.22 networks with enhanced air-interface,” *Pervasive and Mobile Computing*, vol. 9, no. 4, pp. 454–471, September 2013.

[108] J. Tang, S. Misra, and G. Xue, "Joint spectrum allocation and scheduling for fair spectrum sharing in cognitive radio wireless networks," *Computer networks*, vol. 52, no. 11, pp. 2148–2158, April 2008.

[109] M. Vishram, L. C. Tong, and C. Syin, "List multi-coloring based fair channel allocation policy for self coexistence in cognitive radio networks with QoS provisioning," in *IEEE Region 10 Symposium*, July 2014, pp. 99–104.

[110] X. Chen and J. Huang, "Database-assisted distributed spectrum sharing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2349–2361, May 2013.

[111] M. F. Amjad, M. Chatterjee, and C. C. Zou, "Inducing Cooperation for Optimal Coexistence in Cognitive Radio Networks: A Game Theoretic Approach," in *IEEE Military Communications Conference*, November 2014, pp. 955–961.

[112] H. Kamal, M. Coupechoux, and P. Godlewski, "Inter-operator spectrum sharing for cellular networks using game theory," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2009, pp. 425–429.

[113] B. Singh, K. Koufos, O. Tirkkonen, and R. Berry, "Co-primary inter-operator spectrum sharing over a limited spectrum pool using repeated games," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 1494–1499.

[114] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of Interference on Multi-hop Wireless Network Performance," in *IEEE International Conference on Mobile Computing and Networking (MobiCom)*, September 2003, pp. 66–80.

[115] R. Jain, D. M. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC Research Report, 1984.

[116] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, March 2000.

[117] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, August 1980.

[118] 3GPP TS 36.104 v17.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception," Technical Specification, 2021. [Online]. Available: https://www.3gpp.org/DynaReport/36104.htm

[119] C. J. Lanting and A. Rodriguez-Ascaso, "Understanding ICT Standardization: Principles AND Practice," 2019.

[120] FG-NET-2030, "Network 2030 - A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond," International Telecommuications Union, White Paper, May 2019.

List of Publications

Journal Publications

J1 Animesh Kumar, Abhay Karandikar, Gaurang Naik, **Meghna Khaturia**, Shubham Saha, Mahak Arora, and Jaspreet Singh, "Toward enabling broadband for a billion plus population with TV white spaces," *IEEE Communications Magazine*, vol. 54, pp. 28-34, July 2016.

J2 **Meghna Khaturia**, Pranav Jha, and Abhay Karandikar. "Connecting the Unconnected: Towards Frugal 5G Network Architecture and Standardization." *IEEE Communications Standards Magazine*, vol. 4, no. 2, pp. 64-71, June 2020.

Preprint/Submitted/Under Preparation Journal Publications

J3 **Meghna Khaturia**, Pranav Jha, and Abhay Karandikar. "5G-Flow: Flexible and Efficient 5G RAN Architecture Using OpenFlow" Submitted to *Journal of Computer Networks*, 2020.

Patents Filed

P1 **Meghna Khaturia**, Pranav Jha, Abhay Karandikar, "A Wireless Network Architecture", IN201921027947.

Book Chapters

BC1 **Meghna Khaturia**, Sarbani Banerjee Belur, and Abhay Karandikar, "Chapter 5 - TV White Space Technology for Affordable Internet Connectivity," in TV White Space Communications and Networks (R. Stewart, D. Crawford, and A. Stirling, eds.), pp. 83-96, 2018.

Conference Publications

C1 **Meghna Khaturia**, Kumar Appaiah, and Abhay Karandikar. "On Efficient Wireless Backhaul Planning for the "Frugal 5G" Network," in *IEEE WCNC Future Networking Workshop for 5G and Beyond Testbed and Trials*, April 2019.

C2 **Meghna Khaturia**, Jaspreet Singh, Akshay Patil, Sarbani Banerjee Belur, Abhay Karandikar, Prasanna Chaporkar, and Tejashri Ghadge. "Towards Frugal 5G: A Case Study of Palghar Test-bed in India," in *IEEE WCNC Future Networking Workshop for 5G and Beyond Testbed and Trials*, April 2019.

C3 **Meghna Khaturia**, Sweety Suman, Abhay Karandikar, and Prasanna Chaporkar, "Spectrum Sharing for LTE-A Network in TV White Space" in Twenty Fourth *National Conference on Communications (NCC)*, pp. 1-6, February 2018 .

Other Publications

Conference Publications

C4 Sarbani Banerjee Belur, **Meghna Khaturia**, and Nanditha P. Rao. "Community-led Networks for Sustainable Rural Broadband in India: the Case of Gram Marg." *Community Networks: the Internet by the People, for the People.* Association for Progressive Communications, 2017.

C5 **Meghna Khaturia**, Akshatha Nayak, Pranav Jha, Abhay Karandikar, "5G-Serv: Decoupling User Control and Network Control in the 3GPP 5G Network", *Conference on Innovation in Clouds, Internet and Networks (ICIN)*, March 2021.