# QUALITY OF SERVICE AWARE USER ASSOCIATION FOR EFFICIENT NETWORK SLICING IN MULTI-TENANT 5G NETWORKS

By

**SWATHI SUBBIAH JAYANTHI**

A dissertation submitted to the Department of Electrical and Electronic Engineering, Lee Kong Chian Faculty of Engineering and Science Universiti Tunku Abdul Rahman in partial fulfillment of the requirements for the degree of Master of Science

June 2022

## ABSTRACT

## QUALITY OF SERVICE AWARE USER ASSOCIATION FOR EFFICIENT NETWORK SLICING IN MULTI-TENANT 5G NETWORKS

**Swathi Subbiah Jayanthi**

Network slicing is one of the key technologies for fifth generation (5G) communications to overcome hurdles such as increased network density and data traffic due to the growing number of mobile devices. Multi-tenant network slicing allows multiple Mobile Virtual Network Operators (MVNOs) to share the physical network infrastructure, each known as a tenant, for service provisioning. In particular, multi-tenant network slicing for heterogeneous networks (Het-Nets), where small cells co-exist within macro cells, has received significant attention due to the advantages of Het-Nets in boosting network capacity and coverage.

Despite a great amount of research in multi-tenant network slicing for Het-Nets, user association for multi-tenant network slicing in Het-Nets has not well investigated. A few studies have delved into user association for multi-tenant network slicing in Het-Nets, but various performance aspects, namely fairness, quality of service (QoS), user admission, power consumption are not jointly considered. The main objective of this thesis is to study and develop a user association scheme that considers the aforementioned performance aspects for multi-tenant network slicing.

In this thesis, the user association problem for network slicing in multi-tenant 5G Het-Nets is investigated under network slicing scenarios with single and multiple base stations (BSs) per MVNO. In the first part of the thesis, a constrained user association problem that maximizes the weighted sum throughput is formulated under the network slicing scenario with single BS per MVNO. To solve the problem, it is converted into an unconstrained problem and is solved using Genetic Algorithm (GA). Overall, in terms of fairness, QoS, percentage of users associated and the power consumption, the proposed scheme outperforms the existing scheme.

In the second part of this thesis, a user association problem that maximizes the weighted sum throughput is formulated under the network slicing scenario with multiple BSs per MVNO, which has further taken into account intra-slice interference. The problem is again solved using the penalty function approach and GA. The proposed GA based user association scheme outperforms the existing Signal to Interference plus Noise Ratio (SINR) based user association scheme with increased fairness, QoS, number of users associated to the network, and decreased power consumption.

## ACKNOWLEDGEMENT

I would like to express my most sincere gratitude and deepest appreciation to my main supervisor Dr. Lee Ying Loong. I thank him for all his guidance, support, advice, motivation, and drive towards making me think critically to become a better researcher. He steered me in the right direction when he noticed I was so much lagging during the pandemic situation. Also, he never failed to ask my opinion and provided me freedom to share my opinion and difficulties.

My sincere appreciation goes to my co-supervisor Dr. Chang Yoong Choon who was always very supportive, gave me great ideas and helped me realize my mistakes. He helped me a lot while writing my dissertation, by sharing his views on how to improve my writing, how the reviewer will evaluate the dissertation and much more.

I would extend my sincere thanks to my ex-co-supervisor Dr. Simon Lau Boung Yew, for his advice to speed up my work and to complete my studies on time. Although, he resigned from UTAR, he still stays in touch and keep asking me how I am doing and how is my research work. I am grateful to him in all ways.

I would also like to express my sincere gratitude to Dr. Ezra Morris, who helped me to find this research project and with all the admission procedures to get into master's degree in UTAR. On this note, I would also like to thank Dr.

Stella Morris, who have always been a moral support to me and encouraged me whenever I felt low. Thank you, sir and madam.

I thank and acknowledge my dearest friend Ms. Aminath Ibrahim, who was always been with me throughout my journey in Malaysia right from day one and been a pillar of support to share my hard times. My other friends Ms. Nneka and Ms. Hamsalekha who are like sisters to me and been a great support whenever I felt low. I extend my thanks to Ms. Foo Lee Sze for teaching me all the basics of MATLAB and helped me a lot with the coding. Thank you, guys, for all your support.

Last but not the least, I am grateful and thankful to my parents, for all the support and encouragement from my childhood till now. Thank you for giving me all the freedom to pursue my goals and dreams. On this note, I thank my grandpa who is no more, who have brought me up as such a brave girl and my grandma, aunt, uncle and all my well-wishers and friends who never gave me a feel that I am away from home. Thank you all for you love and support.

**APPROVAL SHEET**

This dissertation entitled "**QUALITY OF SERVICE AWARE USER ASSOCIATION FOR EFFICIENT NETWORK SLICING IN MULTI-TENANT 5G NETWORKS**" was prepared by SWATHI SUBBIAH JAYANTHI and submitted as partial fulfillment of the requirements for the degree of Master of Science at Universiti Tunku Abdul Rahman.

Approved by:

(Dr. LEE YING LOONG)
Date: 6 June 2022
Assistant Professor/Supervisor
Department of Electrical and Electronic Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

(Dr. CHANG YOONG CHOON)
Date: 6 June 22
Associate Professor/Co-supervisor
Department of Electrical and Electronic Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

**LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE**

**UNIVERSITI TUNKU ABDUL RAHMAN**

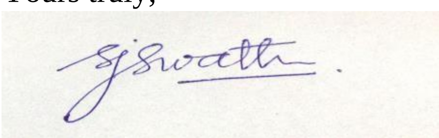
Date: 06/06/2022

## SUBMISSION OF DISSERTATION

It is hereby certified that **Swathi Subbiah Jayanthi** (ID No: **19UEM03863**) has completed this dissertation entitled "QUALITY OF SERVICE AWARE USER ASSOCIATION FOR EFFICIENT NETWORK SLICING IN MULTI-TENANT 5G NETWORKS" under the supervision of Dr. Lee Ying Loong (Supervisor) from the Department of Electrical and Electronic Engineering, Lee Kong Chian Faculty of Engineering and Science, and Dr. Chang Yoong Choon (Co-Supervisor) from the Department of Electrical and Electronic Engineering, Lee Kong Chian Faculty of Engineering and Science.

I understand that University will upload softcopy of my dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,



*(SWATHI SUBBIAH JAYANTHI)*

## DECLARATION

I hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

Name SWATHI SUBBIAH JAYANTHI

Date 06/06/2022

## TABLE OF CONTENTS

## LIST OF TABLES

# LIST OF FIGURES

## LIST OF ABBREVATIONS

| | |
|---|---|
| 1G | First Generation |
| 2G | Second Generation |
| 3G | Third Generation |
| 3GPP | Third-Generation Partnership Project |
| 4G | Fourth Generation |
| 5G | Fifth Generation |
| AI | Artificial Intelligence |
| App | Application |
| API | Application Programming Interface |
| AWGN | Additive White Gaussian Noise |
| BS | Base Station |
| C/U | Control/User |
| CAPEX | Capital Expenditure |
| CeNB | Control Evolved Node B |
| CN | Core Network |
| $CO_2$ | Carbon di Oxide |
| CP | Central Processor |
| C-Plane | Control Plane |
| C-RAN | Cloud Radio Access Network |
| D2D | Device to Device |
| DTX | Discontinuous Transmission |
| E2E | End-to-End |

| | |
|---|---|
| EDF | Earliest Deadline First |
| eMBB | Enhanced Mobile Broadband |
| eNB | Evolved Node B |
| FBEEOS | Forward and Backhaul Link Energy Efficiency Optimization Strategy |
| GA | Genetic Algorithm |
| GPRS | General Packet Radio Service |
| Het-Net | Heterogeneous Network |
| InP | Infrastructure Provider |
| IoT | Internet of Things |
| ITU | International Telecommunication Union |
| JFBLO | Joint Forward and Backhaul Link Optimization |
| KKT | Karush-Kuhn-Tucker |
| LTE | Long Term Evolution |
| M2M | Machine to Machine |
| MANO | Management and Orchestration |
| MBS | Macro Base Station |
| MEERF | Most Energy-Efficient Resource First |
| MIMO | Multiple Input Multiple Output |
| MMS | Multimedia Message Service |
| mMTC | massive Machine Type Communications |
| mmWave | Millimeter Wave |
| MNO | Mobile Network Operator |
| MVNO | Mobile Virtual Network Operator |
| NFV | Network Function Virtualization |

| | |
|---|---|
| NFVI | NFV Infrastructure |
| OFDMA | Orthogonally Frequency Division Multiple Access |
| OPEX | Operational Expenditure |
| OSS/BSS | Operation and Business Support System |
| PA | Power Amplifier |
| PRB | Physical Resource Block |
| QoS | Quality of Service |
| RAN | Radio Access Network |
| RB | Resource Block |
| RF | Radio Frequency |
| RRH | Remote Radio Head |
| RRM | Radio Resource Management |
| SBS | Small Base Station |
| SDN | Software Defined Network |
| SINR | Signal to Interference Plus Noise Ratio |
| SLA | Service Level Agreement |
| SMS | Short Message Service |
| SNR | Signal to Noise Ratio |
| TV | Television |
| UE | User Equipment |
| UeNB | User Evolved Node B |
| U-Plane | User Plane |
| URLLC | Ultra-Reliable Low Latency Communications |
| V2I | Vehicle-to-Infrastructure |

| | |
|---|---|
| V2P | Vehicle-to-Pedestrian |
| V2V | Vehicle-to-Vehicle |
| V2X | Vehicle-to-Everything |
| VNF | Virtual Network Functions |
| Wi-Fi | Wireless Fidelity |

## LIST OF SYMBOLS

| | |
|---|---|
| $B$ | Bandwidth |
| $G$ | Number of generations |
| $i$ | User that belongs to $U$ |
| $k$ | Base station that belongs to $K$ |
| $K$ | Set of Base Stations |
| $m$ | MVNO that belongs to $M$ |
| $M$ | Set of MVNOs |
| $mut$ | Mutation rate |
| $n$ | Number of chromosomes |
| $N0$ | Noise Spectral Density |
| $NF$ | Noise Figure |
| $Sel$ | Selection rate |
| $U$ | Set of users |
| $w1/w2/w3$ | Weighting co-efficient |
| $C_k$ | Fitness of an individual chromosome $k$ |
| $I_{ki}$ | Interference experienced by the user $i$ from the BS $k$ |
| $N_k$ | Number of PRBs allocated to BS $k$ |
| $P_{AWGN}$ | Additive White Gaussian Noise Power |
| $P_{max,j}$ | Maximum transmission power of BS $j$ |
| $P_{max,k}$ | Maximum transmission power of BS $k$ |
| $R_{ki}$ | Throughput of user $i$ on a PRB with BS $k$ |
| $S_k$ | Selection of chromosomes |

| | |
|---|---|
| $V_i$ | Weighting coefficient of the user |
| $W_k$ | Weighting co-efficient of BS |
| $a_{ki}$ | User association variable |
| $g_{ji}$ | Downlink channel gain between the BS *j* to the user *i* |
| $g_{ki}$ | Downlink channel gain between the BS *k* to the user *i* |
| $k_m$ | Base station that belongs to MVNO *m* |
| $p(k)$ | Transmission power of BS *k* allocated to one PRB |
| $\alpha_i$ | Target data rate of user *i* |
| $\tau_{ki}$ | Signal to Noise Ratio |

# CHAPTER 1

# INTRODUCTION

## 1.1 Research Background

As the number of mobile devices, rich multi-media, and cloud data increases day by day in this era, the data traffic and network density has become a major drawback of the past and current generations, (i.e., Third Generation (3G) and Fourth Generation (4G) respectively). Thus, fifth generation (5G) cellular networks is going to emerge to support the growing number of mobile devices and mobile data traffic volumes in the near future. One of the main features for 5G to overcome the challenges posed by the growing number of mobile devices is network sharing. Network sharing involves sharing of cell sites, the common network infrastructure or physical resources like base stations (BSs), backhaul instruments and radio resources between the mobile network operators (MNOs) (Samdanis *et al*., 2016). Network sharing can be realized by creating multiple virtual logical networks known as network slices (Oladejo and Falowo, 2017), with the introduction of two key technologies, namely softwaredefined networking (SDN) and network function virtualization (NFV) (Ordonez-Lucena *et al*., 2017). It is noteworthy that SDN is a technique which is used to design and manage the network in such a way that the network is flexible to respond quickly to the changing business needs and also used to divide the control plane (carrying the signalling traffic) and the data plane (sometimes called as user plane which carries the network user traffic); whereas NFV is an architecture that uses the concept of information technology virtualization to virtualize entire classes of network node function into blocks that may chain together to create communication services. This SDN and NFVenabled network sharing mechanism is called multi-tenant network slicing (LeAnh *et al*., 2017). Multi-tenancy is actually an architecture in which a single instance of a software application serves multiple customers. For multi-tenant 5G network sharing, each tenant refers to an MNO or a service provider that operate over the network slices. Multi-tenancy is also a concept that refers to the logical isolation of shared virtual compute, storage, and network resources (Samdanis *et al*., 2016). The tenant operates over the network slice. It is composed of a specific radio access technology settings and network functions. In network slicing, the MNO acts as an intermediator who acquires the physical resources from the infrastructure provider and bundles the physical resources into virtual resources called slices and sells them to service providers (LeAnh *et al*., 2017). In (Zhao *et al*., 2017; Albonda and Perez-Romero, 2019; Vo *et al*., 2018), the authors proposed various Radio Access Network (RAN) slicing techniques such as Control/User (C/U) plane separation, offline reinforcement learning, centralized and de-centralized approach, slice requirements (such as latency, usage of resources, throughput, etc.,.). The authors of (Elayoubi *et al*., 2019) had discussed the main challenges of RAN slicing by considering the Service Level Agreement (SLA). In (Bahlke *et al*., 2018), the author addressed how to maximize the resource efficiency.

Network slicing was regarded as a virtualization technique that enables multiple logical networks to run on a physical infrastructure of the network. The concept of network slicing allows a shift of paradigm from networks of entities to networks of infrastructures with the necessary capabilities. The network slicing technology optimizes the infrastructure resources that already exists and provides effective savings of capital expenditure (CAPEX) and operational expenditure (OPEX). Network slicing can be divided into core network (CN) slicing and RAN slicing. Allocation of computing resources to create and link virtual CN entities is the central function of CN slicing whereas allocation of BSs, radio resources, and baseband resources is the central function of RAN slicing. RAN slicing is in fact more challenging compared to CN slicing because of the rapidly varying nature of the RAN channel. Energy efficiency is an important key design consideration for multi-tenant 5G networks. Next generation networks may demand high network capacity, high data rate, low latency as well as high energy efficiency. As such, some energy saving schemes had been proposed, as provided in the survey in (Kanwal *et al*., 2017). Generally, the network infrastructure owner can provide the users with ubiquitous connectivity by installing more BSs but that may lead to high power consumption. For the network to be energy-efficient, the power consumption has to be taken into account in 5G network slicing. The energy should be efficiently used by the user to reduce the CAPEX and OPEX (Kanwal *et al*., 2017) and considering the environmental and economical constraints too (Rizvi *et al*., 2017). Another study in (Ge *et al*., 2017) has proved that the improvement of energy efficiency can be done for cellular networks by having the BSs in idle state when they do not share resources to the users. In (Nie *et al*., 2017), the challenges of energy efficiency in networks made of small cells had been studied in the EARTH project to promote improved energy efficiency and proposed a power consumption model at BSs. In (Nie *et al*., 2015), the authors had investigated energy efficiency optimization for the orthogonal frequency division multiple access (OFDMA) small cell network. In (Wang *et al*., 2018), the authors proposed the energy efficient Joint Forward and Backhaul Link Optimization (JFBLO) scheme. Similarly, the authors of (Dai and Yu, 2016) proposed another energy efficient network slicing scheme known as Most Energy-Efficient Resource First (MEERF) scheme. The authors of (Matthisen *et al*., 2018) have presented their work on energy efficient network slicing with the goal of increasing the throughput and energy efficiency. In (Buzzi *et al*., 2016), the authors tried to bridge the gap between the RAN slicing and the energy efficiency. The physical resource block (PRB) which is used to map the physical channels and signals onto resource element in the downlink. One PRB consists of 12 sub-carriers that corresponds to 180 kHz with the 15 kHz subcarrier spacing was introduced in (Liberg and Wikström, 2020).

On the other hand, the user association in 5G network slicing have been discussed in (Ma *et al*., 2018; Ye *et al*., 2013; Zhang *et al*., 2017; Caballero *et al*., 2017). They all suggested various resource allocation methods based on Lotka-Volterra model, load balancing approach and pursue learning methodology.In (Zhang *et al*., 2017) the problem formulated was used to increase the spectral efficiency also. focused on downlink channel association.

## 1.2 Problem Statement

Even though network slicing is promising, management and allocation of network resources for creating network slices and quality of service (QoS) provisioning among tenants are very challenging. Several existing studies on RAN slicing techniques have been conducted to address the challenge (Samdanis *et al*., 2016; Ordonez-Lucena *et al*., 2017; Oladejo and Falowo, 2017; LeAnh *et al*., 2017; Zhao *et al*., 2017; Albonda and Perez-Romero, 2019; Vo *et al*., 2018; Guo and Suarez, 2019; Elayoubi *et al*., 2019; Bahlke *et al*., 2018). In particular, the authors of (Matthisen *et al*., 2018) had discussed the main challenges of RAN slicing is designing and maintaining the slices in an efficient way considering the SLA. However, while resources are shared among tenants in 5G networks, the power consumed by the network could be high if the RAN slicing does not account for the power consumption; which is an important issue for 5G networks. Because of the rapidly growing number of mobile users, the BS power consumption has been increasing. The increase in the power consumption acts as a barrier for the mobile network operators to gain profit in the cellular industry. Since energy efficiency is vital for 5G systems to be sustainable, several research works (Albonda and Perez-Romero, 2019; Vo *et al*., 2018; Nie *et al*., 2015; Wang *et al*., 2018) had been done for addressing the energy efficiency issues for 5G networks. In particular, the authors of (Buzzi *et al*., 2016) tried to bridge the gap between the RAN slicing and the energy efficiency by considering the maximization of throughput by considering transmission power, activation power and quality of service constraints but they did not consider addressing the fairness issue in 5G networks.

On the other hand, various researchers in (Ma *et al*., 2018; Ye *et al*., 2013; Zhang *et al*., 2017; Caballero *et al.*, 2017) had discussed about the various user association schemes in wireless networks. However, these schemes did not focus on fairness, QoS and power consumption. The Signal to Noise Ratio (SNR) based user association scheme was also not efficient in guaranteeing balanced performances. However, the power consumption was not well investigated. From the literature survey, it can be observed and concluded that energy-efficient user association for network slicing and lowering power consumption has not been well-analyzed. In this thesis, the energy efficient user association scheme with decreased power consumption will be addressed.

## 1.3 Significance of Study

The significance of this study is to benefit the telecommunication industry which is aiming in upgrading the 4G technology towards 5G technology. It is because the 5G technology has potential to support millions of devices at ultrafast speed and its key aspects like massive network capacity, low latency, energy efficient usage, etc.,

In this study,

(1) The energy efficient usage is done by minimizing the power consumption of the network. Thus, minimization of power consumption would be helpful for the society by reducing the utility bills, and for the industry by increasing the consumer demands.

(2) The massive network capacity means admitting more users to the network. The user association scheme proposed in this study will maximize the number of users admitted to the network. This also shows the increased fairness of the network.

(3) The low latency is addressed in this study by maximizing the throughput, which eventually maximizes the QoS of the network.

Thus, this study focuses on the key aspects of the 5G networks for the improved performance of the 5G network which would be helpful for the industry to increase its user demands.

## 1.4 Objective

(1) To formulate a user association problem for network slicing in multitenant 5G network.

(2) To develop an artificial intelligence (AI) algorithm to solve the formulated user association problem.

(3) To evaluate the performance metrics such as fairness, quality of service, percentage of users associated in the network, power consumption for the formulated problem.

## 1.5 Thesis Organization

In this thesis, chapter 2 explains in detail about the existing studies, i.e., the literature review. Chapter 3 presents the user association scheme for scenario with only one BS is associated with one Mobile Virtual Network Operator (MVNO). Chapter 4 presents the user association scheme under the scenario where more than one BS is associated with one MVNO. Chapter 5 concludes the thesis by summarizing the Chapter 3 and Chapter 4, and provides several future works.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Evolution of mobile networks

The First Generation (1G) mobile network was first introduced in 1979 in Japan and it relied upon analogue radio systems which meant that users could make phone calls only and could not send or receive text messages. Next, the Second Generation (2G) network was developed in 1991, which ran on digital signals and vastly improved the network security and capacity. Using the 2G network, mobile users were able to send short message service (SMS) and multimedia messaging service (MMS), which was often slow and even not successful sometime. When general packet radio service (GPRS) was introduced in 1997, mobile users were able to send messages and E-mails on the go. Then, the 3G mobile network revolutionized mobile connectivity and capabilities of cell phones. Compared with 2G, 3G connectivity was much faster and able to transmit larger amounts of data.3G communications enabled mobile users were able to make video calls, share files, surf the Internet, play games and watch television (TV) online on their mobile itself. It had become a hub of social connectivity as well. Afterwards, the 4G network emerged as a revolutionary communication technology which was able to provide downlink speed up to 100 Mbps. Users can experience reduced latency, higher voice clarity, easy access to instant messaging services and social media, faster downloads, and quality streaming. Currently, the 5G mobile network is on its way and it is widely anticipated by the mobile industry to provide improved transmission rates and capacity in order to support the new internet of things (IoT) paradigms such as smart cities, smart homes, smart offices and smart cars. Further, experts have predicted that the IoT trend will change not only about how mobile users use our mobile devices but also how mobile users connect our devices to the network (Anon, n.d.).

### 2.2 Key Technologies of 5G wireless system

The key technologies of 5G wireless network are Network Function Virtualization (NFV), Software Defined Network (SDN), Machine-to-Machine (M2M) communication, Massive Multiple Input Multiple Output (MIMO), Device-to-Device (D2D) communication, Millimetre Wave (mm Wave) technology, Network Slicing, Vehicle-to-Everything (V2X) communication, multi-tenant Het-Net.

#### 2.2.1 Massive Multiple-Input Multiple-Output (MIMO)

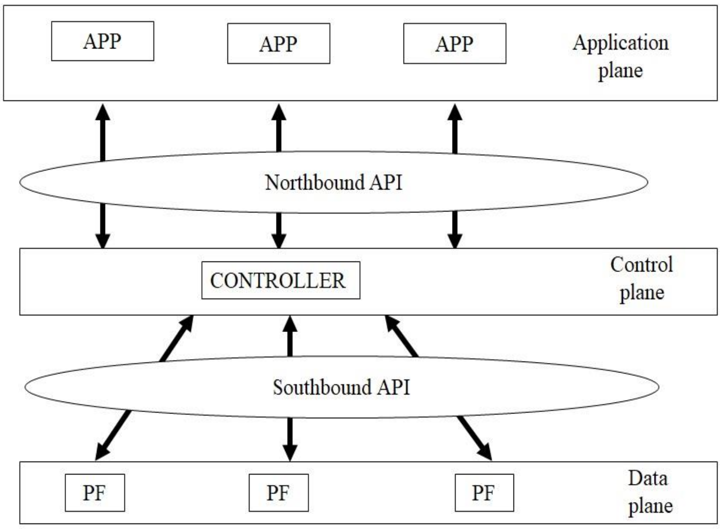
MIMO have been extensively applied on Wireless Fidelity (Wi-Fi), Long Term Evolution (LTE) and so on. Theoretically, MIMO means more antennas, more spectral efficiency and more transmission reliability. In particular, when there are large number of transmit antennas and receiving antennas, the channel capacity of MIMO will increase linearly with minimum value. Massive MIMO can be achieved by some low power and inexpensive components. The advantages are enhancing network coverage and system capacity, exponentially increasing wireless spectrum efficiency (Shi *et al*., 2015).

### 2.2.2 Millimeter Wave (mm Wave)

The frequency range of the mm Wave for 5G network begins at 24 GHz. The mm Wave is the frequency of the 5G network where the frequency is 24GHz or above. The pros of mm Wave are reduced latency rate; increased network throughput because of larger frequency bandwidth range; larger network connection capacity that supports more number of devices and users; reduced data transfer rates; reduced overhead cost which should reduce cost per network connection. However there are cons, they are: smaller cell size; increased number of antennas; higher frequency radio waves (Keith, 2019).

### 2.2.3 Software Defined Network (SDN)

The SDN paradigm makes the optimization of resource allocation and utilization possible in a centralized manner. The management functions in a SDN enabled network is handled by the control plane and it is virtually centralized, separated from the data plane physically, which further aids in enabling high programmability and configurability to the network. Hence, SDN along with network virtualization will play a vital role in 5G cellular network. (Ramantas et al., 2018). Fig. 2.1 depicts the architecture of SDN. It comprises of three layers namely application plane, control plane and data plane (Wu et al., 2015). The application plane is the topmost layer, and it consists of different SDN applications (App). More than one northbound Application Programming Interface (API) drivers and one application logic makes up the application plane. The middle layer is the control plane. The main function of the control plane is to transfer the requirements to the instances of the data plane from the application plane. Finally, the bottom layer is the data plane. The logic network devices of the data plane are exposed through southbound API agents.




| | | |
|---|---|---|
| APP – Applications | API – Application Programming Interface | PF – Packet Forwarding |

Fig. 2.1 Architecture of SDN.

### 2.2.4 Network Function Virtualization (NFV)

The network functions which are implemented in software is separated in an effective manner by NFV. NFV helps to separate hardware devices and network functions, as it is more flexible in terms of adaptive reconfiguration and speed. There are several advantages of NFV, namely, they are centralized networks, virtual versus physical network management, cost optimization, successful network delivery, cloud abstraction and much more. The functions of SDN and NFV when combined can be more advantageous than hardware networks (Sharma *et al.*, 2021). Fig 2.2 illustrates the architecture of NFV. It consists of an Operation and Business Support System (OSS/BSS), a Virtual function layer, and a NFV Infrastructure (NFVI) layer (Yuxiang, 2015). The NFVI consists of collection of cloud-based resources. Physical storage, physical computing and physical networks are virtualized using NFVI, and they are placed into resource pools. The existing telecom service networks are represented by the virtual network layer. Each physical network elements are mapped as virtual network elements or VNF. Virtual computing, virtual storage, virtual network are the virtual elements that require VNF. The operation support layer is OSS/BSS and it should be altered for the purpose of virtualization. The NFV management and orchestration (MANO) domain plays a vital role in differentiating the NFV network from the traditional network. It coordinates and controls all NFVI resources, performs mapping and associates the resources and service networks and executes the procedures of OSS service resources.

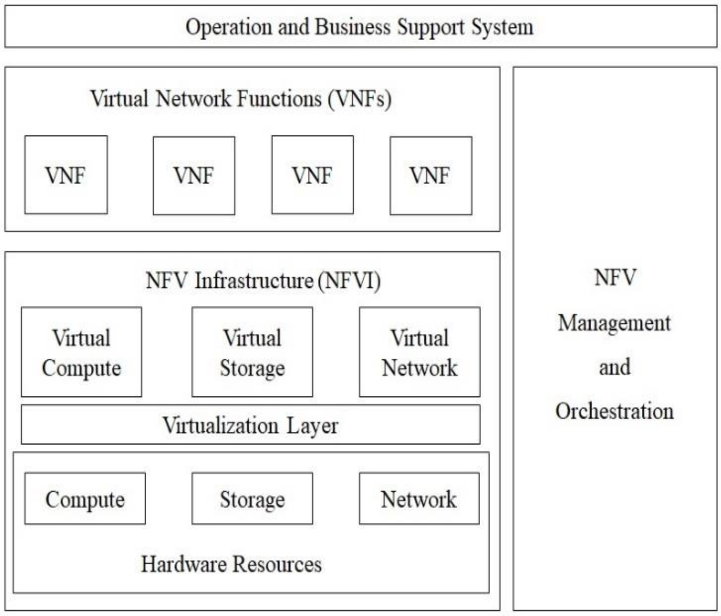



Fig. 2.2 Architecture of NFV.

### 2.2.5 Network Slicing

The network slicing creates end-to-end logical network, starting from the mobile edge, continuing through the RAN mobile transport through the 5G core. Tenants or service specific networks are created by network slices. These tenants are service providers and delivers specific services over the network. Reliability, latency and bandwidth are some of the specific network requirements that the tenants will impose (Keith, 2019).

### 2.2.6 Device-to-Device (D2D) communications

The D2D communication plays a vital role in connecting the user equipment (UE) directly. The D2D communication provides higher data rate, improving coverage and offering peer-to-peer service, spectrum efficiency, and power management was implemented by 5G cellular network. An ISM band which is unlicensed or cellular band which is licensed can provide the resource management and security services from the cellular network can be used by D2D communication. A direct device-to-device connection (also known as peerto-peer connection) removes BS or internet protocol-oriented connection provided by D2D communication.

The different types of D2D communication are network autonomous D2D, network assisted D2D and controlled D2D. In autonomous D2D, the devices create links and communication among them under a fully distributed scenario like ad-hoc, all the device head are able to manage network functions similar to that of self-organized network. In network assisted D2D, the network functions namely security, link management and synchronization, are all supported by infrastructure. In network controlled D2D, all devices are accessible for data communication if and only if the network is completely centralized. Further, D2D communication can also be classified under in-band mode and Out-band mode. In in-band D2D, the resources are reused, and specific resource shares the same spectrum of cellular and D2D devices. Whereas in out-band D2D, different spectrums are used for cellular network in the field of industry, medical and scientific bands. (Sharma *et al*., 2021).

### 2.2.7 Machine-to-Machine (M2M) Communication

M2M communication is the set up using the sensors, actuators, objects and machines and which works without human assistance. M2M communication is used for transmitting a sensed data which is small and with the time constraints. The network capacity must be large to manage the massive amount of concurrent connections in a wide coverage area. Data offloading and data aggregation are used to amplify communication and energy efficiency in M2M communication. For the automatic operations of data generation, processing and transferring, intelligent machines are provided by M2M communication (Sharma *et al*., 2021).

### 2.2.8 Vehicle-to-Everything (V2X) Communication

The V2X communication is a vehicular technology strict requirements of increased data rates, reliability and low latency with the advantages of security services, autonomous vehicles and traffic information systems. It includes vehicle-to-pedestrian (V2P), vehicle-to-infrastructure (V2I), and vehicle-to-vehicle (V2V) communication. As D2D provides short end-to-end (E2E) latency and long transmission range, it is one of the enabling technology for V2X communication. The features of V2X communication are: (1) Reliability – $10^{-5}$; (2) E2E delay – 10 to 100ms; (3) Data rate – 10 to 40 Mbps; (4) Positioning accuracy – 30 cm (Sharma *et al*., 2021).

### 2.2.9 Multi-Tenant Heterogeneous Network

The small cell BSs of 5G cellular Het-Net are unevenly distributed and has an uneven service area. Generally, the small cell BSs are scattered and clustered in the macro cell network to supplement and support the communication in the hot spot area and the edge area. The network topology and the cell coverage distribution are the major differences between the 5G HetNet and the traditional macro cellular network where the BSs are evenly distributed usually on a hexagon grid and the service area corresponds to the hexagon on the grid (Ai *et al.*, 2021).

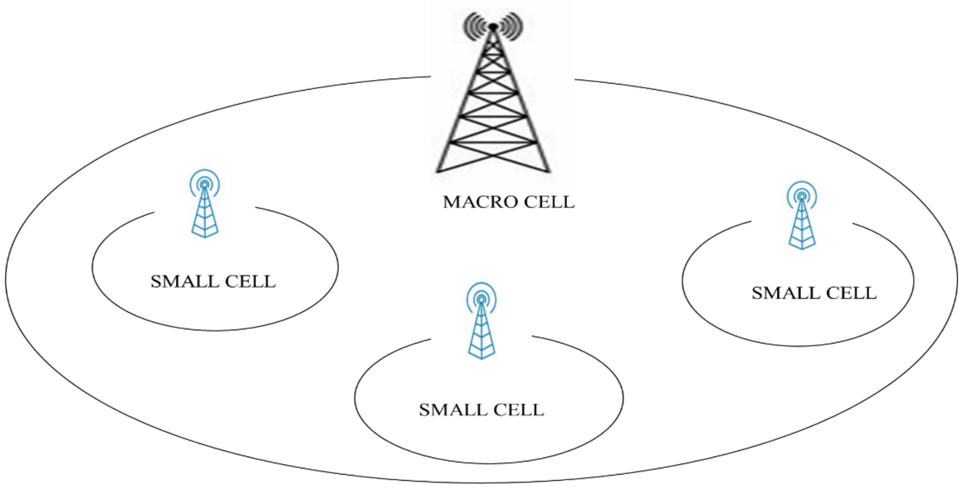



Fig. 2.3 Heterogeneous Network.

Radio resources are transferred in terms of physical radio Resource Blocks (RBs) among many heterogeneous BSs using the multi-tenancy technique. The sharing of physical RAN infrastructure with several mobile virtual network operators (MVNO) is called as multi-tenancy. Fig. 2.3 shows the simple structure of a Het-Net which comprises both macro cell and many small cells. Fig. 2.4 illustrates the multi-tenant Het-Net. The network where various types of small BSs are deployed under the macro-BS and shares RAN infrastructure with the MVNOs are called as multi-tenant Het-Net.

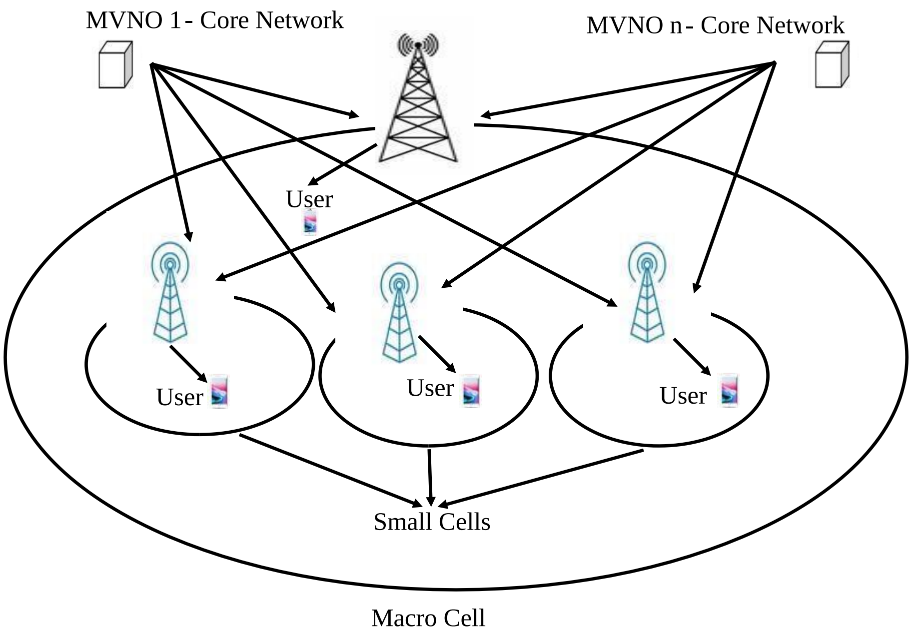



Fig. 2.4 Multi-Tenant Heterogeneous Network

This thesis focuses on network slicing for a multi-tenant Het-Net. In particular, user association, that is the process of the network determining the appropriate serving BS for each mobile user, is investigated for 5G network slicing in multi-tenant Het-Nets.

### 2.3 Network Sharing and Slicing

As the number of mobile devices, rich multi-media, and cloud data have been gradually increasing in the recent times, the data traffic and network density has become a major drawback of the past and current generations, (i.e., 3G and 4G respectively) with less CAPEX and OPEX. Thus, 5G cellular networks have emerged to support the growing number of mobile devices and mobile data traffic volumes. One of the main features for 5G to overcome the challenges posed by the growing number of mobile devices is network sharing. Network sharing involves sharing of cell sites, the common network infrastructure or physical resources, such as BSs, backhaul instruments that comprises the intermediate connections between the core network and the small sub-networks at the edge of the network, and radio resources between the MNOs. The mobile network operators share the network infrastructure to offer the users with faster network rollouts and lower-cost services. Network sharing helps shorten the network investment payback period in rural areas, whereas in urban areas, network sharing allows for avoiding the regulatory issues that are causing the difficult and lengthy site-acquisition processes. The evolution of network sharing to multi-tenancy is dependent upon the software-based capabilities and virtualization methods that are introduced to the third-generation partnership project (3GPP)-based networks, which greatly impacts its standardization roadmap. These capabilities expand on the concept of network slicing for providing specific communication services with a specific way of handling control/user plane for this service. Network slicing was implemented by assigning not only network capacity, but also virtual network functions (VNFs), computational resources, per slice customized control/user-plane divides, shared network functions across multiple slices (Samdanis *et al.*, 2016).

Network sharing can be realized with the introduction of two key technologies, namely SDN and NFV. These technologies can enable flexibility, programmability and modularity necessary to build many virtual networks socalled network slices, each customized to a particular use case on top of a common network for network softwarization. The key aspects that are necessary for realizing the network slicing concept are resources, virtualization, orchestration, and isolation. A resource is a manageable unit that can be used to provide a service and is specified by a collection of attributes or capabilities. The two resources are network functions and infrastructure resources. The effective sharing of resources among the slices is enabled by network slicing is called virtualization. The incorporation of virtualization into the networking sector facilitates the development of modern business models, with new players and distinct business functions. Consider a scenario in which there are three types of actors: Infrastructure Provider (InP) who owns and manages the provided constituent resources and physical networks; tenant who leases the virtual resources from one or more InP as virtual networks, through which they manage, realize and provide network service to the user; and end user who receives the services provided by the tenant. Orchestration in network slicing is to coordinate the disparate network processes for creating, managing, and delivering services. Isolation is a critical prerequisite for operating parallel slices on a single shared underlying substrate. The main challenges of network slicing of 5G are performance issues in a shared infrastructure, management and orchestration issues, security and privacy (Ordonez-Lucena *et al.*, 2017). SDN is a technique which is used to design and manage the network in such a way that the network is flexible to respond quickly to the changing business needs and also used to divide the control plane (carries the signalling traffic) and the data plane (sometimes called as user plane which carries the network user traffic). On the other hand, NFV is an architecture that uses the concept of information technology virtualization to virtualize entire classes of network node functions into blocks that may be chained together to create communication services. This

SDN and NFV-enabled network sharing mechanism allows for enabling multi-tenant network slicing.

With SDN and NFV, network slicing can be realized by creating multiple virtual logical networks known as network slices (LeAnh *et al*., 2017). In order to meet a specific use case, each slice includes radio resources, core network resources and backbone. On an as-a-service basis, a physical network can be separated into numerous virtual networks, resulting in the creation of a virtual operator network. RAN slicing is the process of dividing RANs into many virtual networks, which facilitates a paradigm change from an entity-based network to a capability-based network. A fundamental concept behind multitenant RAN slicing is to share a physical RAN infrastructure with many MVNOs. The MVNO is addressed as the tenant in this network architecture paradigm, while the InP owns the physical infrastructure. The physical resources of the RAN, including frequencies, antennas, BS, computing entities and power are virtualized into slices. And the resources are shared among the MVNOs.

According to the International Telecommunication Union (ITU), “The SLA is a formal agreement between two or more entities that is reached after a negotiating activity with the scope to assess service characteristics, responsibilities and priorities of every part” (Anon, 2002). The SLA is a formal agreement between a service provider and a tenant, or between service providers, that specifies the level of service delivered. Because the criteria and characteristics of different service types in legacy telecommunication networks are almost identical, most service provider-tenant SLAs contain the same metrics. In slice-based 5G networks, however, each slice requires its own SLA, with distinct components, metrics, and structure from the SLAs of other slices in the same network (Habibi *et al*., 2018).

## 2.4 Resource Allocation for 5G Network Slicing

A number of studies have been conducted on various network slicing techniques in (Oladejo and Falowo, 2017; LeAnh *et al*., 2017; Zhao *et al*., 2017) and various multi-tenant network slicing (Zhao *et al*., 2017; Albonda and PerezRomero, 2019; Vo *et al*., 2018; Guo and Suarez, 2019; Elayoubi *et al*., 2019). In (Oladejo and Falowo, 2017), the authors proposed two resource allocation schemes for 5G network slicing. They are the centralized resource allocation scheme and hierarchical resource allocation scheme. The InP performs resource allocation to the various users of each MVNO under the centralized resource management strategy and makes resource allocation inefficient and difficult. In hierarchical schemes, however, different entities namely MVNOs and InPs are engaged in resource allocation. These various entities are classified into various levels of hierarchy and share duties of allocation of resource duties based on their classifications (Oladejo and Falowo, 2017).

The authors of (LeAnh *et al*., 2017) proposed a slice isolation technique for the uplink of a virtualized cellular network, in which BSs and chunk-based radio resources controlled by separate InPs segregate virtualized resources or slices. This isolation enables customization of one slice and it will not interfere with other slices since they are individually defined. To find suboptimal decision on slice and transmit power allocations, the authors also proposed a distributed algorithm based on Lagrangian relaxation. The problem was solved in two stages: power allocation and slice allocation, which are accomplished by modifying the sequence of primal and dual variables. The optimal power was obtained from Karush-Kuhn-Tucker (KKT) conditions, and the slice allocation is solved centralized using the Hungarian method. To avoid the need for global details, the authors proposed a distributed algorithm based on the idea of the matching game. It was shown that this algorithm converges to a suboptimal solution (LeAnh *et al*., 2017).

In (Zhao *et al*., 2017), the authors proposed a technique for RAN slicing based on C/U plane separation for air interfaces that took into account network functions, logical channels, physical channels, physical signals, and user status. Meanwhile, it separated the control plane from the user plane and further dividing the evolved NodeB (eNB) (which is a macro cell BS of cellular networks, that was responsible for resource management, scheduling, modulation, demodulation) into control evolved NodeB (CeNB) and user evolved NodeB (UeNB), where the CeNB was responsible for providing large coverage with low frequencies and UeNB was responsible for high-speed transmissions with high frequencies in a small coverage range. To manage and assign resources to the collection of RAN slices, the authors designed a centralised controller. There are three layers in the RAN slicing scheme based upon C/U plane separation, and they were: the virtual controller, the virtual CeNBs/UeNBs and the remote radio heads (RRHs) pool.

The RAN slicing problem had also been investigated in (Albonda and Perez-Romero, 2019), to provide the generic services of 5G networks, which were the V2X and enhanced mobile broadband (eMBB) services. With respect to the investigated problem, the authors proposed an offline reinforcement learning based on RAN slicing scheme. A low-complexity heuristic algorithm was developed to allocate radio resources to various slices. The algorithm aims in increasing the utilization of resources. The authors formulated an optimization problem to keep up a V2X and an eMBB slice on the same infrastructure of RAN to estimate the resources allocated to each slice with the goal of optimizing usage of radio resource while meeting the unique needs of each slice. The developed model considered downlink, uplink and side-link transmission. As an improved approach for determining the appropriate resource split between the two slices, a novel technique based on offline Qlearning and soft-max decision-making was suggested. The slice controller continuously interacted with an environment model in this approach to understand the optimal policy. The used Q-learning algorithm monitored every alternative "actions" (i.e., dividing of resources) that the system might undergo an exploration-exploitation process. Based on the results of the off-line Qlearning method, a low-complexity heuristic strategy was given for fine-tuning resource assignment and obtaining additional gains in resource utilization. The suggested approach's performance was tested using comprehensive simulations to demonstrate its capacity to execute effective resource allocation across slices in terms of latency, resource consumption, outage probability and attainable data rate. The simulation results demonstrated the efficiency of the suggested strategies under various system conditions.

In (Vo *et al.*, 2018), the authors formulated a bi-convex RAN slicing problem. The formulated problem takes into accounts about the interdependencies between coordination of slices and slicing of resources that share the resources. The two algorithms namely centralized and de-centralized approach was proposed to solve the formulated bi-convex problem. Firstly, it is the centralized approach, that could be held up by orchestration tools and 5G network virtualization. Next, it is the de-centralized or distributed approach is where in-order to attain a solution, all slices co-ordinate. The authors also mentioned that although, the two approaches does not guarantee global convergence, a global optimal solution can be achieved by simulation results by considering a simple scenario having only two tenants.

A novel framework had been introduced in (Guo and Suarez, 2019) for enabling the RAN slicing with various slice requirements in terms of latency, throughput, usage of resources. These requirements are expressed as number of resources needed per interval deadline. The slices are arranged as per the Earliest Deadline First (EDF) principle. For the first time, the concept of EDF scheduling, which was first utilized in real-time operating systems and is used in scheduling of RAN slice. The relevant system architecture is also provided, which includes self-organizing function modules and slice admission control. The approach that was proposed by the authors might satisfy both hard and soft requirements of latency for radio access, as well as assuring the bit rate via mapping of proper throughput with resource needs. If the capacity of system is sufficient, it can handle an any number of slices. The authors also considered the requirements of the RAN slicing scheme from the Radio Resource Management (RRM) perspective. Those requirements are: (1) slices can be generated, deleted, or updated dynamically; (2) different slices will have varying geographical coverage; (3) the slices may have various service instances, that require either guaranteed or non-guaranteed performance; (4) different slices may have extremely different performance requirements in terms of latency, throughput, dependability, availability, and customization of these settings is required; (5) the guaranteed performance of a slice is unaffected by the circumstances of other slices with equal or lower priority; (6) a tenant must be able to apply unique scheduling and admission policies inside its slice; (7) from the standpoint of a network operator, the total resource usage must be increased.

The authors of (Elayoubi *et al*., 2019) had discussed the main challenges of RAN slicing were maintaining and designing the slices among the infrastructure which is shared in an effective way in such a way that the SLA is considered. The challenges also included the slice isolation. The other challenges were resource allocation, SLA monitoring and multi-tenant network slicing. The authors also mentioned that the slicing can be treated as a method of simplifying and optimizing network and infrastructure sharing across operators. The proposed work is carried out by identifying many slice granularity choices where each slice is allocated to family of service, smartphones, internet of things, technical requirements, group services that comes under the same family in terms of requirements, vertical customer, business customer and technical requirements. Given the diversity of their requirements, these services cannot be adequately controlled by a single slice. The fourth choice was One Slice Defined Per Business Customer and Technical Requirements where this option defines a slice for a certain automotive customer for on-board entertainment, a slice for the same customer for autonomous driving, another slice for another automotive customer for autonomous driving, and so on. This option is consistent with the 3GPP standard's description of "a network slice instance". The authors of (Elayoubi *et al*., 2019) have presented their work on energy efficient network slicing by considering the network slicing aware joint bandwidth and power allocation, with the goal of increasing the throughput and energy efficiency.

To increase the efficiency of resources of a wireless communication network, the authors of (Bahlke *et al*., 2018) had proposed an optimization technique by jointly optimizing resource distribution among network slices. The available time-frequency resources in dense and heterogeneous cellular 5G networks were divided orthogonally among multiple slices that were supplied by the cells. The authors proposed a scheme that optimizes distribution of resources between network slices, cell allocation to employ on separate slices, and user to cell allocation simultaneously. The authors also showed how to include bandwidth efficiency and Signal to Interference plus Noise Ratio (SINR)-requirements of transmission methods devoted to different services into the network optimization process. They also investigated the impact of cell design with various pools of resources which are orthogonal on efficiency of system resources. As a starting point for optimizing resource efficiency, the authors employed the cutting-edge strategy of complete frequency reuse and allocating demand points to the cell with the strongest signal.

Many studies had been done on investigating energy efficiency for cellular networks. In (Kanwal *et al*., 2017), the authors explained about the various energy saving schemes. Over that, there are two major classification of power consumption at BS: Static power consumption and Dynamic power consumption. Static power consumption mostly remains constant and belongs to the hardware and the BS needs to provide power for operations. It can be improved by energy efficient designs and subsequent intelligent deployments. Whereas the dynamic power consumption relies upon the resource utilization by the BS, and it can be affected by the BS’s transmission operation. The authors mainly focused on the latter. During off-peak time periods, proper activation and deactivation of BS's transceivers, also known as Discontinuous Transmission (DTX), can minimize dynamic or communicational power usage. A number of energy saving schemes were proposed in the existing work, and by analyzing the pros and cons of the energy saving schemes, the authors had concluded that the scheme called ‘Dynamic Distance Aware Based Energy Saving Scheme’ will save more power as it involves the process of activation and deactivation of the BS when the BS is not in use. Also, most of these research had been done on achieving energy-efficient cellular networks by adjusting the transmission power from the base-station to the users or vice versa. The user would expect pervasive network connections in next-generation wireless networks, in addition to the overwhelming need for high data rates and network capacity. This type of networking method would yield increased network capacity and energy efficiency, as well as improved network coverage.

In (Rizvi *et al*., 2017), a multi-tier network had been considered, which consists of three scenarios, i.e., macro cells, macro cells overlaid by pico-cells, macro cell overlaid by pico-cells and femto-cells. All these tiers have different properties including cell size, transmitting power and density. It is evident that the size of the pico-cell and femto-cell are smaller than the macro cell, as they are over laid under the macro cell. The density of tier 1, which consists of macro cells is $\lambda_{macro}$, tier 2 in which macro cells over laid by pico-cells is $\lambda_{pico} = 2\lambda_{macro}$, tier 3 in which macro cells are over laid by pico-cells and femto-cells is $\lambda_{pico} = 2\lambda_{macro}$ , $\lambda_{femto} = 10\lambda_{macro}$ . The network is modelled using stochastic geometry. The data energy efficiency was investigated by deploying more cells in the form of multi-tier Het-Nets. It is noteworthy that the study considered the scenarios where two-tier networks involve macro cells and picocells and three-tier networks involve macrocells, pico-cells and femto-cells. Both networks are subjected to energy efficiency assessments. The effect of the number of Pico and Femto BSs on energy efficiency was also investigated in depth. From the study, the authors found that the 5G Het-Net's energy efficiency can be boosted to a higher extent if small-cell BSs, such as pico and femto cells, are grown in a reasonable manner, according to statistical and numerical research. Furthermore, efficient deployment of the correct number of BS can result in increased energy efficiency. Moreover, the analysis of energy efficiency was only based on the power transmitted by BSs within the tier under review. If effective power transmission and adaptive power control mechanisms are utilized, energy efficiency is predicted to improve even further.

In (Ge *et al*., 2017), the authors explained about the various types of power consumption at BSs considering the functional and architectural structure of BSs, namely additional power, computation power and transmission power. Computation power is the amount of energy utilized by baseband units, which includes digital signal processing operations, BS management and control operations, and communication between the core network and BSs. All of these activities are carried out by software on semiconductor devices. Transmission power, is the amount of energy consumed by power amplifiers (PAs) and radio frequency (RF) chains to modify wireless signals, i.e. signal transformation between baseband and wireless radio signals. Additionally, the power consumed at feeders is included in the transmission power. Additional power defined by the study in (Ge *et al*., 2017) is the power consumed by BSs, excluding transmission and calculation power, such as the power required to keep BSs running. The study indicated that the challenges in the 5G small cell network to face the role of computational power are: (1) The influence of 5G network topologies on computational power in 5G small cell networks; (2) With the technologies such as millimeter wave transmission and huge MIMO the computation power is optimized; (3) finding the trade-off between computation and transmission power. The study in (Ge *et al*., 2017) had proved that energy efficiency can be maximized for cellular networks by having the BSs in idle state when they do not share resources to the users.

In (Nie *et al*., 2017), under user data rate constraints, this research attempted to improve the system energy efficiency of OFDMA small cell networks. The authors considered the effects of both backhaul and forward links while formulating the optimization problem of system's energy efficiency. The ideal solution to the system energy efficiency problem necessitated the simultaneous consideration of scheduling, transmission power optimization, and backhaul connection data rate regulation. Then using the analysis of the specified problem of system's energy efficiency the authors computed and determined the lower and upper bounds of the ideal energy efficiency of the system. Then, to approach the established system energy efficiency constraints, forward and backhaul link energy efficiency optimization strategy (FBEEOS) is suggested to decrease signalling overheads and computation complexity. The challenges of energy efficiency in networks made of small cells had been studied in the EARTH project.

In (Nie *et al*., 2015), the energy efficiency of the system of the OFDMA small cell network was improved by considering the user equipment (UE)'s data rate and backhaul effect. The goal of this article (Buzzi *et al*., 2016) is to maximize the OFDMA small cell network's energy efficiency. In contrast to prior energy efficiency research that focused on BSs, the impact of backhaul lines was investigated using a buffer-less data flow model. It was suggested to use a JFBLO methodology to repeatedly improve energy efficiency of the system by optimizing backhaul data rate and power emitted. In JFBLO, the energy efficiency of the system was shown to be non-decreasing and convergent. The authors proposed the JFBLO scheme to improve the system energy efficiency by iteratively optimizing the forward link (i.e., the logical link connecting the BS and the users) and the backhaul links because the emission power affects the performance of forward link, which consequently affects the backhaul link's power consumption, which influenced the energy efficiency of the system. The simulation findings showed that, while accounting for a tiny percentage of overall system power consumption, emission power had a significant impact on forward link performance, which in turn affected backhaul link power consumption and system energy efficiency.

In (Wang *et al*., 2018), wireless network slicing was studied as a potential technique for next-generation networks that enables mobile consumers to receive personalized on-demand services. The authors had considered a wireless network slicing scheduling policy aimed at maximizing the network's energy efficiency, measured as the ratio of the long-run average throughput of user requests to the long-run average power consumption. This created a problem with exceptionally high computational complexity, making it impossible to use traditional optimization approaches directly. The authors of (Matthisen *et al*., 2018) proposed the MEERF scheme, which was a scalable priority policy to maximize the energy efficiency of the network, by always prioritizing the communication nodes with the highest effective energy efficiency. In the particular scenario applicable for a local wireless environment with a dense user population and exponentially distributed service time need, MEERF was shown to be asymptotically optimum. Extensive simulations indicated MEERF's tolerance to diverse service time distributions. By comparing MEERF to benchmark policies in a larger network with possibly geographically scattered users and infrastructures, the authors can quantify its efficacy. MEERF outperforms the benchmark policies, according to the findings.

The cloud radio access network's (C-RAN) energy efficiency was investigated (Dai and Yu, 2016), with an emphasis on two essential and distinct downlink transmission techniques, they are data-sharing and compression techniques. The backhaul lines linking the BSs and central processor (CP) were used to convey user messages in the data-sharing approach - messages of each users are delivered to several BSs; then the BSs generate the beamforming vectors locally and then jointly delivers the messages from BSs to the users. The messages of users are pre-coded at the CP, and the CP sends a compressed version of the analogue beam-formed signals to the BSs for combined transmission in the compression method. The energy efficiency of the two techniques were compared in this study by constructing an optimization problem to minimize total network power consumption while meeting target rate of the user as limitations, where overall network power comprised of load dependent backhaul power, BS activation power and BS transmission power. The paper's major result is that, in a wireless cellular network, that C-RAN greatly increases the range of possible user data rates. Both the schemes are greatly enhanced downlink C-RAN energy efficiency.

In (Matthisen *et al.*, 2018), the authors mentioned that network slicing can enable 5G network operators to serve different tenants with varying service needs. In terms of throughput and energy efficiency, this article addressed network slicing informed optimum resource allocation. For a sliced radio access network with per-slice zero forcing beamforming, the authors designed a heterogeneous QoS framework and jointly optimize power and bandwidth allocation across slices and users. Two distinct strategies based on the utility profile and scalarization techniques along with generalized fractional programming were used to find the Pareto optimum solutions of this multi objective optimization problem. Numerical findings demonstrated the benefits of assigning bandwidth and transmission power together, as well as how slice specific QoS requirements affect throughput and overall energy efficiency. In (Matthisen *et al*., 2018), the authors tried to bridge the gap between the RAN slicing and the energy efficiency in throughput, transmission power, activation power and QoS constraints.

In (Buzzi *et al*., 2016), the authors provides an overview about the energy-efficient wireless communication. Energy efficiency have become a significant feature in designing and operating the wireless communication system. Indeed, the network system aims only to optimize the latency, data-rate and throughput for more than a century. In this decade, energy efficiency also accounts in operating wireless communication systems, considering the economic, operational and environmental aspects as well. As energy efficiency have emerged as a key aspect, the paradigm shift from optimizing the throughput to optimizing the energy efficiency have commenced. In (Buzzi *et al*., 2016), the authors also tried to bridge the gap between the RAN slicing and the energy efficiency in throughput, transmission power, activation power and QoS constraints.

In (Xu et al., 2021), the authors have compared the Het-Net with homogeneous network which helps increases the QoS of users and reuse of resources, by developing small cells into macro cell coverage. Efficient resource allocation algorithms play a vital role in minimizing the interference and to achieve spectrum sharing. They have proposed a control-based resource allocation algorithm which has dynamic characteristics. They have also mentioned that this algorithm will be very helpful in future trends of Het-Nets.

The PRBs were introduced in (Liberg and Wikström, 2020). One PRB consists of 12 sub-carriers that corresponds to 180 kHz with the 15 kHz subcarrier spacing. The concept of PRB is used for the mapping of physical channels and signals onto resource element, in the downlink. Whereas, in uplink the resource units are used for the mapping of physical channels and signals onto resource elements. Each PRB pair is assigned to two separate time slots. In the time domain, each slot is defined by a resource grid consisting of 7 orthogonal frequency division multiplexing (OFDM) symbols and 12 subcarriers in the frequency domain. Each subcarrier has a width of 15 kHz.

## 2.5 User Association for 5G Network Slicing

Before data transmission begins, a user association mechanism is required to identify whether the user is linked with a certain BS. The function of user association in improving network load balancing, spectrum efficiency, and energy efficiency is crucial. In (Ma *et al.*, 2018), the authors mentioned that virtualization technology seems to be a good way to improve resource usage and interference management in Het-Nets by abstracting radio resources. The wireless network infrastructure may be detached from the services it offers using virtualization technologies, allowing different services to utilize the same infrastructure. The Het-Net's physical radio resources held by InPs are extracted and divided into virtual radio resources, forming a virtual radio resource pool in wireless virtualization. The virtual radio resources are leased by MVNOs and assigned to users. Because Het-Net's architecture and physical resources have been abstracted and divided into virtual resources, several effective methods, such as interference control and load balancing, have become easier to be implemented. Furthermore, it is conceivable for many MVNOs to cohabit on the single InP and share infrastructural and radio resources, maximizing resource usage while lowering CAPEX and OPEX. The authors proposed a resource management scheme in which different physical networks' radio resources were virtualized into a virtual resource pool, and MVNOs competed with each other for the virtual resources from the pool to offer services to consumers. The utility function which is aggregated was incorporated into the Lotka-Volterra model (biological model) and assessed both fairness, utility and system dynamics. In the proposed virtual radio resource management scheme, a virtual resource allocation algorithm based on the Lotka-Volterra model was also constructed. The suggested approach was capable of quickly capturing the system's time dynamics. It would also be beneficial to investigate the dynamics of the users over a period of time.  They have also proven that the proposed algorithm had achieved a better trade-off between the throughput and fairness than the existing techniques.

In (Ye *et al*., 2013), the authors proposed a theoretical optimization approach for load-balancing problem in which the resource allocation and cell association were jointly considered. Then by assuming that the users are able to be served with multiple BS, the joint utility maximization problem was decoupled into two sub-problems and it can be solved individually on two perspectives i.e., user's side and BS's side. This approach serves as a benchmark for the upper bound on the network utility to be achieved. The study mainly focused on downlink channel association. Although it is applicable to the uplink channel association, the usage of uplink power regulation, which alters the interference based on the relationship, complicates things.

In (Zhang *et al*., 2017), multi-tenant BSs are introduced as a new network architecture in which a single BS is leased to multiple operators. Multiple tenants can transmit from a single BS using their own spectrum resources. In this article, the authors proposed a user to multi-tenant BS association scheme that is implemented at a centralized controller. Each mobile user can connect or associate with more than one multi-tenant BSs. The authors began by formulating an optimization problem to maximize the spectral efficiency of an macro base station (MBS) network using UEs-to-MBSs relationships as optimization variables. To solve this problem, they conceived an optimization approach based on the pursue learning methodology, in which each optimization variable is represented by a learning automaton. Each automaton in the automata system computes the UE to MBS association in parallel at the centralized controller. In comparison to other current systems, this results in exceptionally low computing complexity and good scalability. Furthermore, as compared to a single tenant BS network, modelling findings reveal that the MBS network has much superior spectrum efficiency.

In (Caballero *et al*., 2017), RAN resources are sliced by numerous tenants, was addressed by the authors. They considered resource allocation among tenants based upon the weighted proportionally fair objective, that aimed to achieve a required degree of fairness across the network slices of different tenants. Several fundamental features are established, including the Pareto optimality of user association to BSs, the fair distribution of BS resources, and the utility benefits and capacity savings arising from dynamic resource sharing between slices. Several significant features such as fair allocation of resources to BSs arise as a result of dynamic resource sharing between slices and pareto optimality of user to BS association. Then, the practical and algorithmic challenges were addressed. The key contribution had been to demonstrate that, despite its complexity, workable systems that scale to huge networks and follow network load fluctuations can be designed. They showed that the goal is NP hard, making an exact solution impractical. The authors provided a realistic solution based on this technique that has low computational loads, information, and handoff overheads. This study showed that the proposed strategy was near optimal and delivered significant benefits to both renters (in terms of capacity savings) and end-users (in terms of cost savings and improved performance).

In (Khalili et al., 2020), the authors proposed a joint user association, antenna selection, carrier allocation and power control in uplink of a Het-Net to maximize the data rate of the small cell users. The problem that they have considered is a non-convex and mixed integer non-linear integer programming problem. They have divided the formulated problem into two as: (i) user association, antenna selection and carrier allocation, and (ii) power control. They solved the decomposed problem iteratively using the majorization-minimization theory and augumented Lagrange method. This algorithm is referred to as Joint Power Control and Scheduling algorithm (J-PCS). They obtained a locally optimal solution for each sub-problem created.

In (Dinh et al., 2021), the authors studied about the architecture or design of sustainable Deep-Q-Network (DQN). The vital role of the DQN is to optimize the user association in a Beyond 5G (B5G) Sub-6GHz and mmWave integrated network. In order to increase the long-term sum rate and the QoS of various applications, they have proposed a $\varepsilon$-greedy policy at each users' DQN execution. They have also provided a detailed study of energy consumed by each user device. They have also mentioned that their numerical results shows the decrease in energy cost required by default user DQN.

## 2.6 Critical Analysis of Current Advances in 5G Network Slicing

Even though network slicing is promising, management and allocation of network resources for creating network slices and quality of service (QoS) provisioning among tenants are very challenging. Several existing studies on RAN slicing techniques have been conducted to address the challenge (Samdanis *et al*., 2016; Ordonez-Lucena *et al*., 2017; Oladejo and Falowo, 2017; LeAnh *et al*., 2017; Zhao et al., 2017; Albonda and Perez-Romero, 2019; Vo *et al*., 2018; Guo and Suarez, 2019; Bahlke *et al*., 2018). In particular, the authors of (Elayoubi *et al*., 2019) had discussed the main challenges of RAN slicing is designing and managing/maintaining the slices in an effective way among the shared infrastructure by considering the SLA. However, while resources are shared among tenants in 5G networks, the power consumed by the network could be high if the RAN slicing does not account for the power consumption, which is an important issue for 5G networks.

Because of the rapidly growing number of mobile users, the BS power consumption has been increasing. The increase in the power consumption acts as a barrier for the mobile network operators to gain profit in the cellular industry. Since energy efficiency is vital for 5G systems to be sustainable, several research works (Kanwal *et al*., 2017; Rizvi *et al*., 2017; Ge *et al*., 2017; Nie *et al*., 2017; Nie *et al*., 2015; Wang *et al*., 2018; Dai and Yu, 2016; Matthisen *et al*., 2018) had been done for addressing the energy efficiency issues for 5G networks. In particular, the authors of (Buzzi *et al*., 2016) tried to bridge the gap between the RAN slicing and the energy efficiency by considering the maximization of throughput by considering transmission power, activation power and quality of service constraints but they did not consider to address about the fairness in resource allocation and backhaul power consumption in 5G networks. From the literature survey, it can be observed and concluded that energy-efficient network slicing has not been well-investigated. This thesis aims to address the energy efficiency issues in network slicing for multi-tenant 5G network by minimizing the power consumption of the network. The energy efficiency issues are important to be addressed because energy efficiency influences both the economic (increase in CAPEX and OPEX) and environmental (due to high emission power, the Carbon di-oxide ($CO_2$) level in the atmosphere increases, which further leads to global warming) constraints.

On the other hand, the authors of (Ma *et al*., 2018; Ye *et al*., 2013; Zhang *et al*., 2017; Caballero *et al*., 2017; Khalili *et al.*, 2020; Dinh *et al.*, 2021), focused on user association schemes in wireless network. Although there are various techniques, they were not so promising in terms of QoS, fairness index and power consumption. Traditional user association schemes such as maximum SNR-based user association were not efficient as high SNRs do not guarantee balanced performance between the aforementioned aspects. However, power consumption was not well investigated in the user association context so far. In this thesis, user association with efficient power consumption will be investigated.

## 2.7 Conclusion

Although many researches have been carried out on two directions: resource allocation and user association. The existing resource allocation schemes such as EDF principle, C/U plane separation, etc., have been done by many researchers, however they did not take into account energy efficiency indepth. On the other hand, user association in multi-tenant Het-Net was not studied well so far. Maximum SNR-based user association was done previously but the energy efficient way of user association was not addressed. Similarly, on the other hand, many researches have been carried on various energy efficient schemes and energy saving schemes, where the authors did not provide a clear figure about the energy efficient user association schemes. Other researchers are needed to give a solution for energy efficient user association scheme.

# CHAPTER 3

# USER ASSOCIATION FOR NETWORK SLICING IN MULTI-TENANTHETEROGENEOUS NETWORKS WITH SINGLE BS PER MVNO

## 3.1 Introduction

As discussed in the previous chapter, energy-efficient user association in a multi-tenant network has not been investigated thus far. As a result, the goal of this chapter is to address the problems with the existing systems by posing them as a mathematical problem. This chapter gives the detailed description about the system model and the problem formulation. The latter is arranged as follows: Section 3.2 describes the multi-tenant 5G Het-Net system model; Section 3.3 briefs about the problem statement; Section 3.4 explains about the problem formulation.

## 3.2 System Model

A Het-Net with one MBS and multiple small base stations (SBSs) is considered in this section. The SBSs can be of femtocells, microcells and picocells. These MVNOs provide services to users such as connectivity and communication over the deployed infrastructure that includes the physical resources such as BSs. The SBSs are deployed within the coverage of the MBS.

The system model of the Het-Net is shown is Fig.3.1, which depicts that only one BS is deployed under each MVNO (Tran and Le, 2020). It is noteworthy that there is no specific set of users assigned to each MVNO. All the users can access the network by associating with any MVNO via its BSs. However, each MVNO does not share the  same BS. The MBS and SBSs have different sizes of coverage areas. For example, the MBS covers a larger area of coverage within the mobile network, whereas the SBSs individually cover smaller areas of coverage. The areas that the SBSs cover are usually those where the received signal strength from the MBS is low. These areas include shopping malls, university, apartments, etc. The MVNO share the network resources within the coverage area to serve the users in the network.

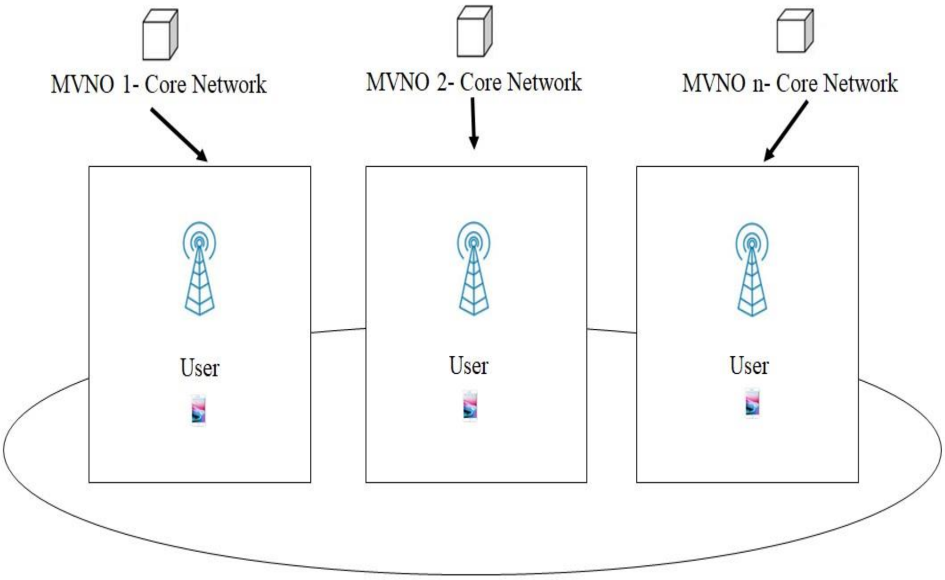



Fig.3.1 Multi-Tenant Heterogeneous Network Having One BS Associated with One MVNO

### 3.3 Problem Statement

Even though there are various approaches proposed by various researchers in the aspect of resource allocation and user association in a multitenant Het-Net, those approaches were not promising in terms of fairness, QoS, power consumption and energy efficiency. Thus, in this chapter, the identified research gap will be formulated as a mathematical problem and the performance metrics such as fairness, QoS, power consumption and the percentage of users admitted to the network are estimated and compared with the existing SNR based user association scheme.

### 3.4 Problem Formulation

Let $K$ and $U$ be the set of BSs (one MBS and several SBS) and users respectively. It is also assumed that each MVNO associates with one of the BSs to provide their services. Next, the total number of PRBs are divided orthogonally among the BSs and each BS receives $N_k$ number of PRBs.

Let us consider $a_{ki}$ as the user association variable, where $a_{ki}$ is equal to 1 if the user *i* connects with BS *k*, 0 otherwise. Each user *i* can achieve a throughput on a PRB by the BS *k*, which can be evaluated using Shannon's capacity formula. The following is the throughput equation of user *i* on a PRB with BS *k*.

$$R_{ki} = Blog_2(1 + \tau_{ki}) \tag{3.1}$$

where $B$ is the spectrum bandwidth of a PRB, and

$$\tau_{ki} = \frac{P_{max,k}\, g_{ki}}{P_{AWGN}} \tag{3.2}$$

is the SNR between user $i$ and BS $k$. In (3.2), the $P_{max,k}$ is the maximum transmission power of BS $k$, $g_{ki}$ is the downlink channel gain between the BS $k$ to the user $i$, and $P_{AWGN}$ is the additive white Gaussian noise (AWGN) power. It is assumed that the user association is carried out on a coarse-grained time scale. Furthermore, it is assumed that the channel gain of all PRBs are averaged over the user association period and each user receives only one PRB. Moreover, $\propto_i$ is defined as the target data rate of user $i$.

To this end, this study aims to increase the number of users associated with the BS deployed under that MVNO in such a way that the throughput is maximized, while taking in account the fairness of the network QoS and the network power consumption. Therefore, the user association problem is formulated as:

$$max_{(a_{ki})} \sum_{k\in K} \sum_{i\in U} W_k V_{ki} a_{ki} R_{ki} \tag{3.3}$$

subject to,

$$\sum_{k\in K} a_{ki} \leq 1, \forall\, i\ \in U \tag{3.3a}$$

$$\sum_{i \in U} a_{ki} \leq N_k, \ \forall k \in K \tag{3.3b}$$

$$a_{ki} \in \{0, 1\}, \ \forall i \in U, k \in K \tag{3.3c}$$

where, $a_{ki}$ is the user association variable, i.e.,

$$a_{ki} = \begin{cases} 1 \text{ if user } i \text{ is associated with BS } k \\ 0 \text{ otherwise} \end{cases}$$

In constraint (3.3a), one user is allowed to associate with only one BS. Constraint (3.3b) ensures that the total number of PRBs should be equal to or greater than the number of users associated to the BS, as it is assumed that one user only receives only one PRB. $W_k$ is the weighting co-efficient of BS $k$, which helps to optimize the objective function by prioritizing the BS which has the lower transmission power, and further helps to increase the energy efficiency. $V_{ki}$ is the weighting coefficient of the user $i$ corresponding to BS $k$, such that $0 \leq V_{ki} \leq 1$, which helps to optimize the objective function by prioritizing the user which has the higher data rate requirement $\alpha_i$ of the user $i$, and thus further increases the throughput $R_{ki}$ of the Het-Net. To this end, the following is set:

$$W_k = \frac{1}{P(k)} \tag{3.4}$$

where, $P(k) = \frac{P_{max,k}}{N_k}$, is the transmission power of BS $k$ allocated to one PRB, while,

$$V_{ki} = \frac{R_{ki} - \alpha_i}{\alpha_i} \tag{3.5}$$

It is worth noting that, although QoS is a multi-attribute performance metric, where the attributes include latency and packet loss, this thesis focuses on the data rate as the QoS requirement, because guaranteeing sufficient data rates would also ensure low latency and low packet loss rates.

## 3.5 Proposed Algorithm and Solution

Genetic Algorithm (GA) is inspired from the natural evolution theory of Charles Darwin. It is a search heuristic algorithm that imitates the process of natural selection, where the fittest individuals are selected for reproduction in order to produce off-springs in the next generation. In fact, GA is designed to solve optimization problems with discrete binary variables e.g., 0-1 integer programming problems. As the problem in (3.3) is a 0-1 integer programming problem with a huge search space, which cannot be solved easily, this makes GA a suitable approach for the problem. GA has been used to solve constrained and unconstrained optimization problems based on a principle similar to biological evolution and natural selection processes. The reason of adopting GA is that it can efficiently solve optimization problems with huge search spaces and large numbers of variables.

GA is an optimization algorithm which is used to find the optimal solution or solutions for a given computational problem by maximizing or minimizing an objective function. The objective function is termed as fitness function which is to be optimized by the algorithm. The fitness function is the crucial part of the algorithm. The GA generally begins with the initial population (i.e., the first generation) which is formed by randomly chosen chromosomes. There are three main operators in GA (Carr, 2014).

(1) Selection: The main part of GA is selection. The strategy behind selection is: “the better fitted an individual, the larger the probability of its survival and mating”. The direct way of implementing this strategy is roulette wheel selection (Lipowski and Lipowska, 2012). The equation that describes the probability of selecting a chromosome is:

$$S_k = \frac{\sum_{k=1}^{N} C_k}{C_k} \tag{3.6}$$

In the above equation, $C_k$ is the fitness of an individual chromosome $k$, $C_k > 0$, $N$ is the total number of chromosomes, $k = (0,1,2, \ldots, n)$, $k$ is the BS.

(2) Crossover: The crossover is implemented after selection. The purpose of the crossover is to create new off-springs by exchanging the genes or bits of the chromosomes or strings selected from the mating pool. The crossover point is selected randomly along the length of the string and the bits on the right side of the strings selected are swapped or exchanged to form two new off-springs. The following Fig.3.2 shows the illustration of single point crossover.

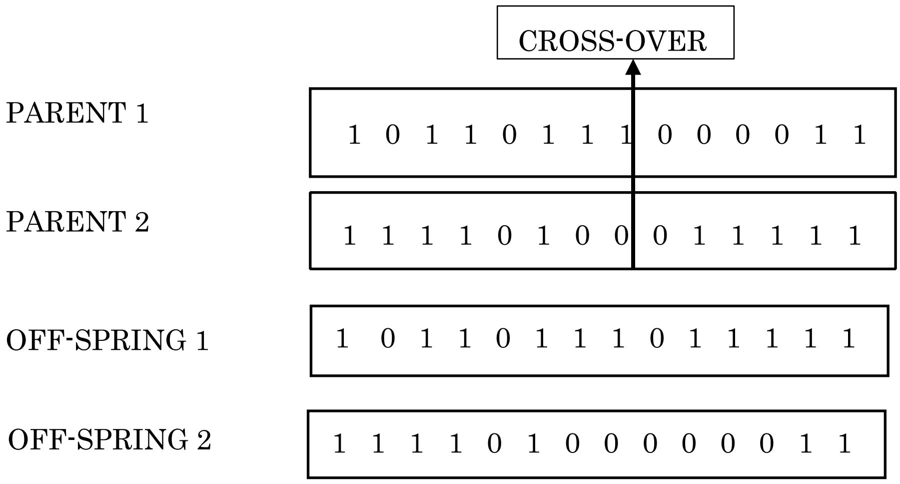



Fig.3.2 Illustration of Crossover Operator

(3) Mutation: Mutation is a GA operator performed to the off-springs formed as a result of the crossover process. With the small mutation rate *mut*, the mutation operator is applied to the new off-springs. The binary bit is changed from 0 to 1 or vice versa using the mutation operator. In the single point mutation, the mutation site (i.e., the bit to be mutated) is selected randomly along the length of the string and the binary digit at the mutation site is changed from 0 to 1 or vice versa. The main purposes of mutation are to (1) maintain the diversity in the population; (2) prevent the early loss of essential genetic material at a specific location; (3) create a string (design point) in the vicinity of the existing string, completing a local search around the present solution. The following Fig.3.3 shows the illustration of single bit mutation.

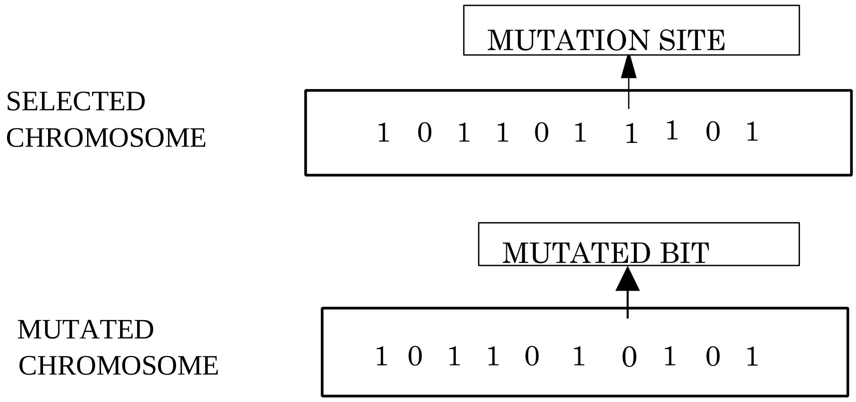



Fig.3.3 Illustration of Mutation Operator

In GA, the penalty functions are the concepts used to handle the constraints, which penalize the unfeasible solution by lowering the fitness value and in proposition to the value or degree of violation of constraint. Thus, in GA, the constraints are converted into penalty functions, thus converting the constrained problem into unconstrained problem. The penalty function term is added to the objective function to convert the constrained problem into unconstrained problem. The constraint in equation (3.3a) and (3.3b) are converted into second and third term of the unconstrained problem in equation (3.7) given below. Thus, the constrained problem in (3.3) can be rewritten into unconstrained problem as

$$max_{(a_{ki})} w_1 \sum_{k \in K} \sum_{i \in U} w_k v_{ki}\, a_{ki}\, R_{ki} \; + \; w_2 \frac{\sum_{i \in U} \langle 1 - \sum_{k \in K} a_{ki} \rangle}{|U|} + \; w_3 \frac{\sum_{k \in K} \langle N_k - \sum_{i \in U} a_{ki} \rangle}{|K|} \qquad (3.7)$$

where, $w_1$, $w_2$, $w_3$ are the weighting co-efficient for each term in the objective function and their values are obtained by performing many trials and selecting the values that provides the best network performance. These values are mentioned in Table 3.2, and

$$\langle Z \rangle = \begin{cases} Z \; if \; Z < 0 \\ 0 \; otherwise \end{cases} \tag{3.8}$$

The problem in (3.7) is solved by using the GA. The equation (3.7) is considered as the objective function as it has the maximization function included in order to maximize the user association, which is the vital objective of this thesis. Also, the constrained problem in equation (3.3) has been changed as unconstrained problem in (3.7), which is easier to solve. The equation (3.7) is also the fitness function as it has the weighting coefficients ($w_1$, $w_2$, $w_3$) which ranges from 0 to 1 and helps to optimize the given problem. The values of the weighting coefficients can be assigned in such a way that the best desired performance can be achieved. The BSs in the network are considered as chromosomes in the context. The initial population is first created as a matrix with the dimension $K * U$. This is the user association matrix. This matrix is converted into a chromosome (i.e., a vector form) by the equation below:

$$\text{chromosome}\,(:,n) = \text{round}(\text{rand}(K * U,1)) \tag{3.9}$$

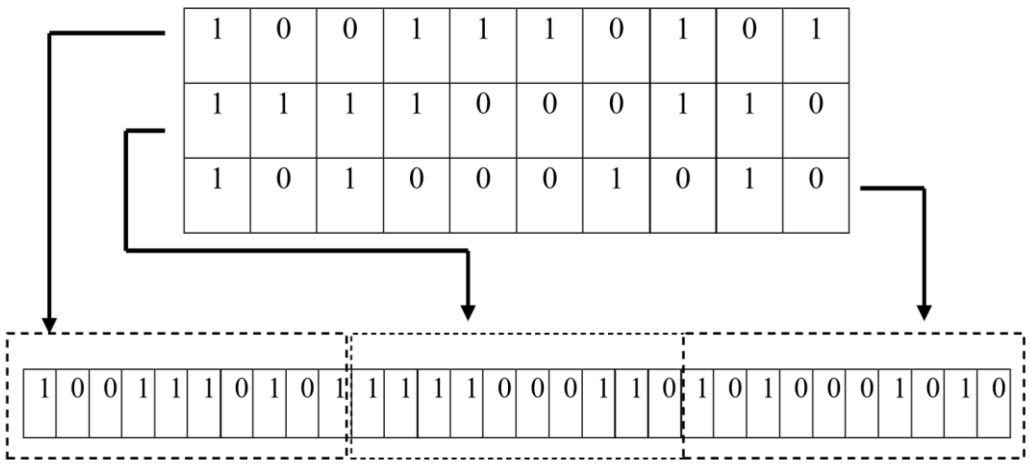



Fig.3.4 Illustration of User Association Matrix to a Chromosome

The above equation 3.9 has been illustrated in Fig.3.4. The user association matrix has been converted into a chromosome as a vector form. It has been done to find the fitness of the chromosome at each iteration.

In the proposed algorithm, the Roulette Wheel Selection is used. The number of chromosomes is denoted by *n*. Once the chromosomes are created, the best chromosomes are selected and that is preserved for the selection process, using roulette wheel selection, the parent chromosomes are selected. Among the parent chromosomes, two chromosomes are randomly selected for crossover to produce new off springs. Then, mutation happens, where the bit is changed. Table 3.1 shows the summary of the proposed GA-based user association scheme.

TABLE 3.1: PROPOSED GENETIC ALGORITHM FOR THE USER ASSOCIATION IN A MULTI-TENANT HETEROGENEOUS NETWORK SLICING.

| | |
|---|---|
| **Step 1** | Generate an initial population matrix 'a' of size ($K*U$), where $K$ and $U$ are the total number of BS and users in the network respectively. To be noted, the number of BSs and number of MVNOs are the same in the network. |
| **Step 2** | Set the selection rate *sel* = 0.5, mutation rate *mut* = 0.1, number of generations $G$ =100, number of chromosomes $n$ = 100. Each chromosome is a candidate solution for (3.3). |
| **Step 3** | Compute the fitness function (objective function) using equation (3.7) for each chromosome. |
| **Step 4** | Select the best chromosome (BS) from the initial population according to the selection rate. The genetic algorithm begins with the selection process using Roulette Wheel selection. |
| **Step 5** | The selected chromosomes are the parent chromosomes. |
| **Step 6** | Single point crossover is performed over the selected chromosomes. The crossover point is calculated, the bits from the crossover point to the last bit of the parent chromosomes are swapped between themselves to produce two new off-springs.<br><br>A bit flip mutation is performed over the off-spring chromosomes. The bit which is to be mutated is chosen randomly and that bit is flipped (one to zero or vice versa). |

| | |
|---|---|
| **Step 7** | Compute the fitness function (objective function) using equation (3.7) for each chromosome. |
| **Step 8** | Sort out the best fit chromosome by calculating the fitness of chromosomes by comparing the fitness of chromosome of previous generation and the current generation during each generation. |
| **Step 9** | A new generation population (user association) matrix of size ($K*U$) is developed from each chromosome. |
| **Step 10** | Repeat step 4 to step 9 until the maximum number of generations has been reached. |

### 3.6 Results and Discussion

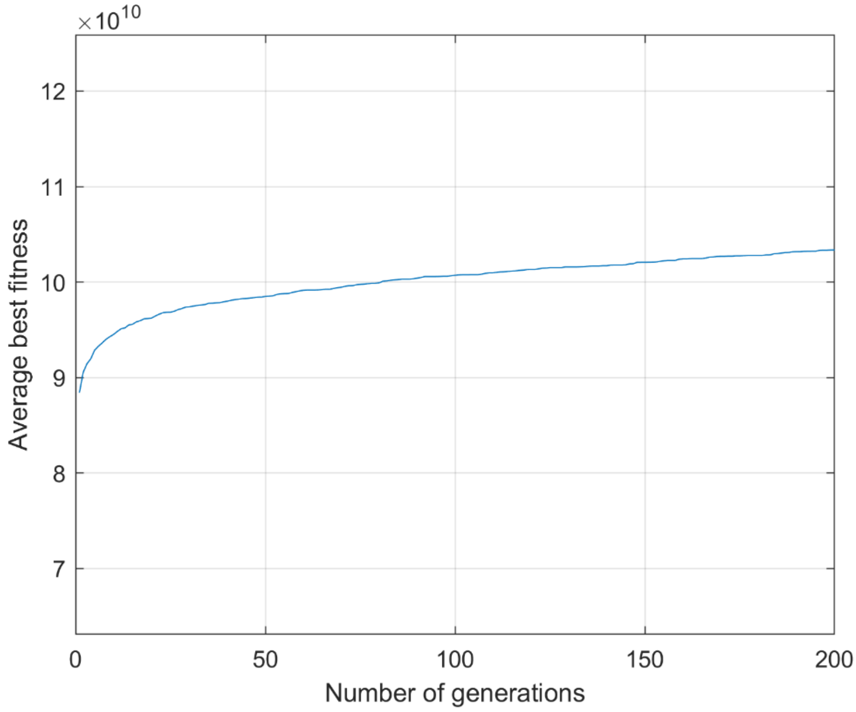
For performance comparison, a traditional maximum SNR based user association scheme, denoted by 'Existing SNR based algorithm' (Ye *et al.*, 2013) is considered. In this algorithm, the user is associated with the BS who provides the highest received SNR for the user. The following Table 3.2 shows the parameter settings for the simulation. Also, the channel model is pathloss channel. The SBS-to-UE pathloss is given as 140.7+36.7log($d$), and MBS-to-UE pathloss is given as 128.1+37.6log($d$), where $d$ is the distance between the user and the BS and is given in unit km. The weighting co-efficient values are calculated using trial and error method whereby numerous different values of the weighting co-efficient values are tested and the values that provide the best network performance are selected.

TABLE 3.2: SIMULATION PARAMETERS

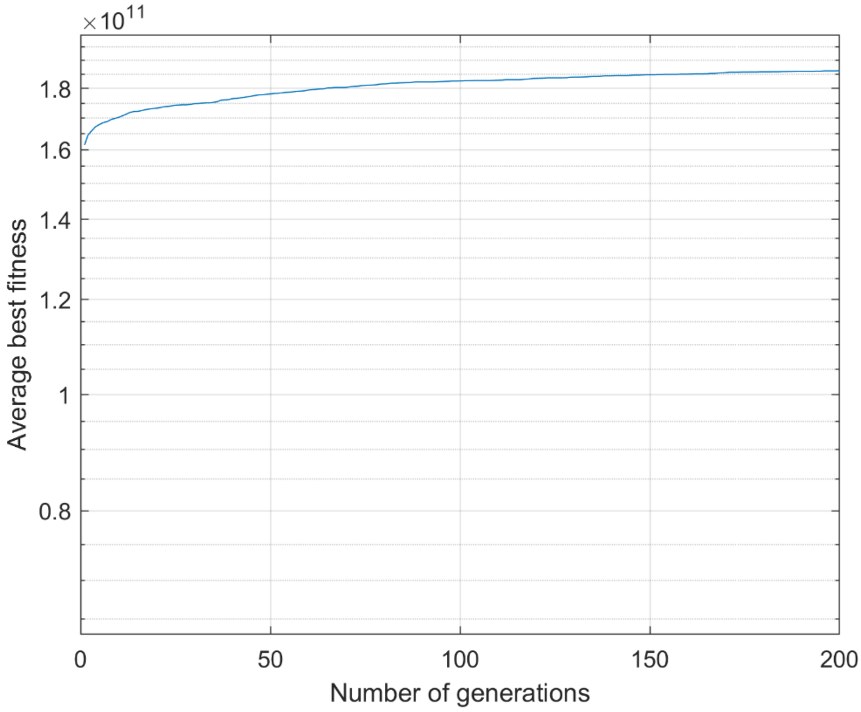
| **Parameters** | **Symbols** | **Values** |
|---|---|---|
| No. of MVNO | $\lvert M \rvert$ | 4 |
| No. of BS | $\lvert K \rvert$ | 4 |
| No. of Users | $\lvert U \rvert$ | 20,40,60,80 |
| Bandwidth of each PRB | $B$ | 180 kHz |
| Noise figure | $NF$ | 9 dB |
| Noise Spectral Density | $N0$ | -174 dBm |
| Data Rate Requirement | $\propto_i$ | [100, 300] kb/s |
| Selection Rate | *sel* | 0.5 |
| Mutation Rate | *mut* | 0.1 |
| No. of chromosomes | $n$ | 100 |
| No. of generations | $G$ | 100 |
| Weighting Co-efficient | $w_1/w_2/w_3$ | 0.9/0.001/0.099 |

The performance evaluations are done as follows.

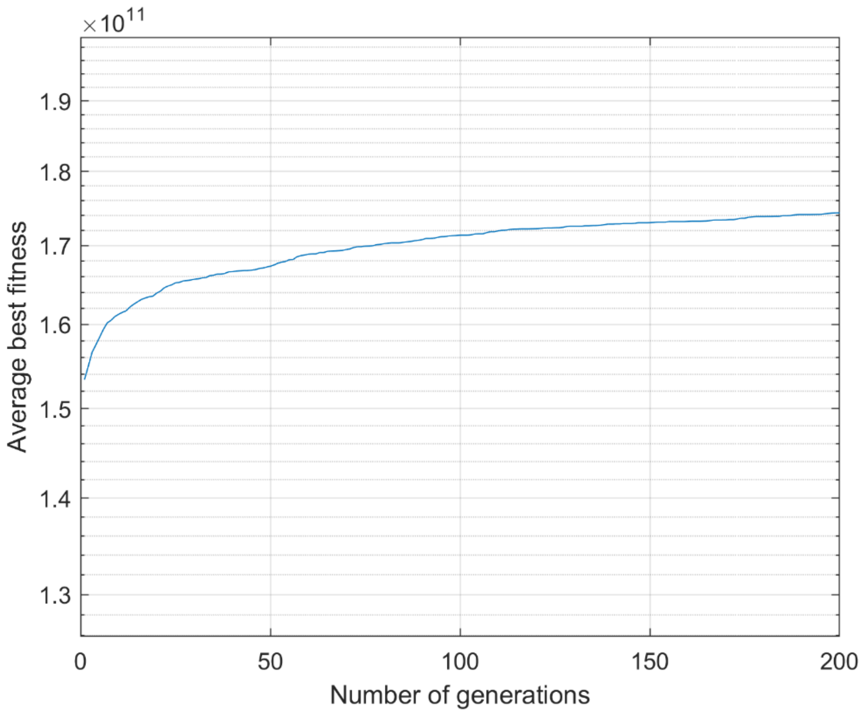
Fig.3.5 shows the convergence performance of the proposed GA based user association scheme for various number of user such as (a) 20 (b) 40 (c) 60 (d) 80 users. It is observed that for all the four scenarios, the convergence of the proposed scheme commences at the 50th generation on average over 100 runs. As a next step, the performances of the proposed algorithm will be evaluated and compared with the existing SNR based user association.



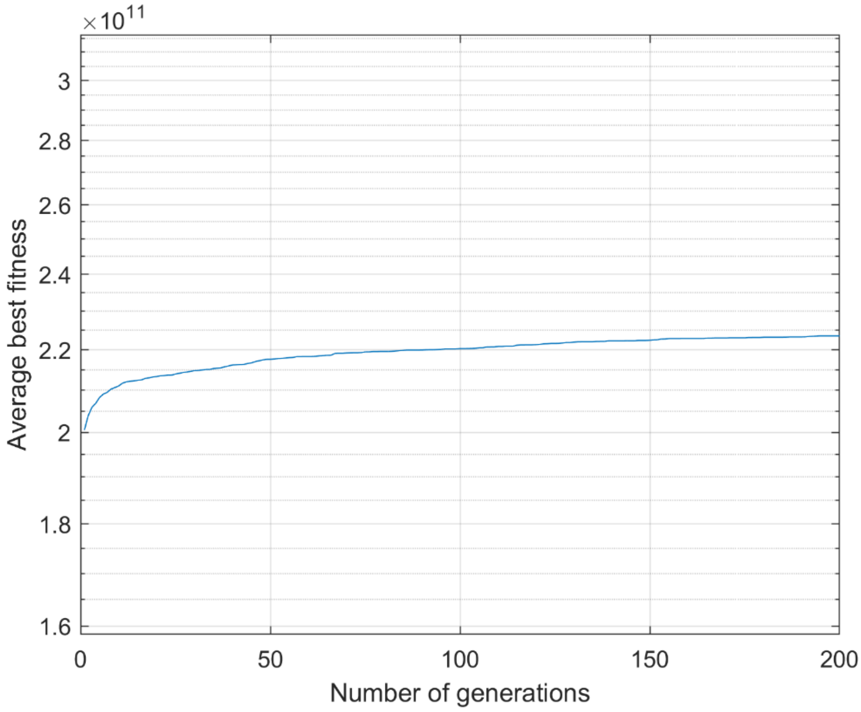

(a)


×10^11
1.8
1.6
1.4
1.2
1
0.8
Average best fitness
0
50
100
150
200
Number of generations

(b)


×10^11
1.9
1.8
1.7
1.6
1.5
1.4
1.3
Average best fitness
0
50
100
150
200
Number of generations

(c)




(d)

Fig 3.5 Illustration of GA convergence for scenarios with (a) 20, (b) 40, (c) 60 and (d) 80 users.

### 3.6.1 Fairness Index

Jain's fairness index (Lee *et al.*, 2015) is used to calculate the fairness index of the network using the following.

$$Fairness\ Index\ =\ \frac{(ATM'1'+\cdots+ATM'n')^2}{M*\left(ATM'1'^2+\cdots+ATM'n'^2\right)} \qquad (3.10)$$

Where, *ATM 'n'* is the throughput achieved for MVNO *n*.

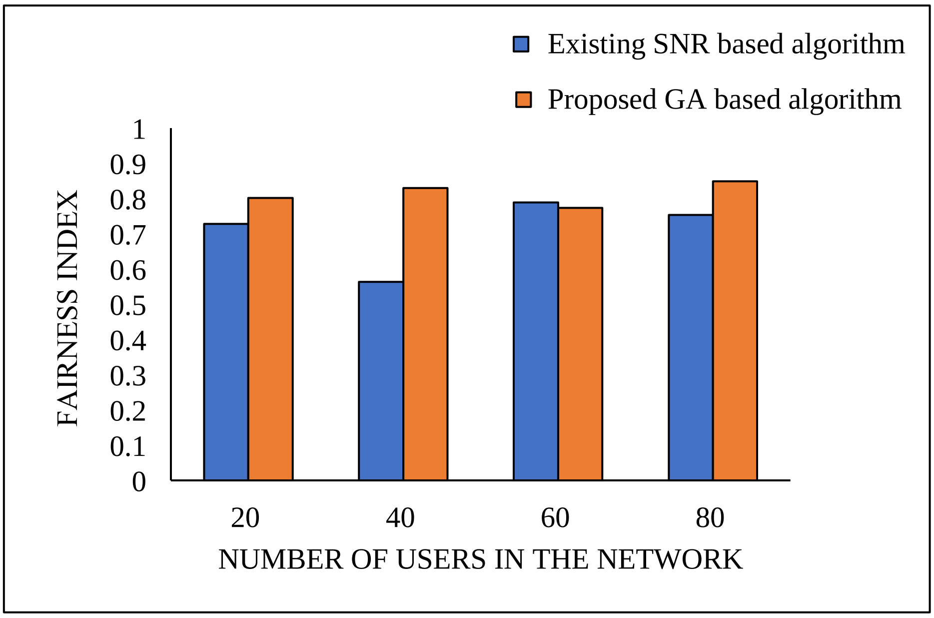



Fig.3.6 Comparison of Fairness Index Between the Existing Algorithm and Proposed Algorithm

In Fig.3.6, the fairness performance of the Het-Net is shown. It is observed from the Fig.3.6 that the fairness performance at the instance of 60 users slightly drops down for the proposed scheme. It is due to the contention between the users to associate with the given number of BSs in order to provide better performance to the users. At the remaining instances, the performance of both schemes slightly varies. This shows that the proposed scheme shows to outperform the existing SNR-based scheme, because the former aims to fulfil the QoS requirements of each user belonging to its MVNO.

### 3.6.2 Quality of Service

QoS is evaluated as the percentage of users that achieve the required data rate divided by the total number of users $|U|$.

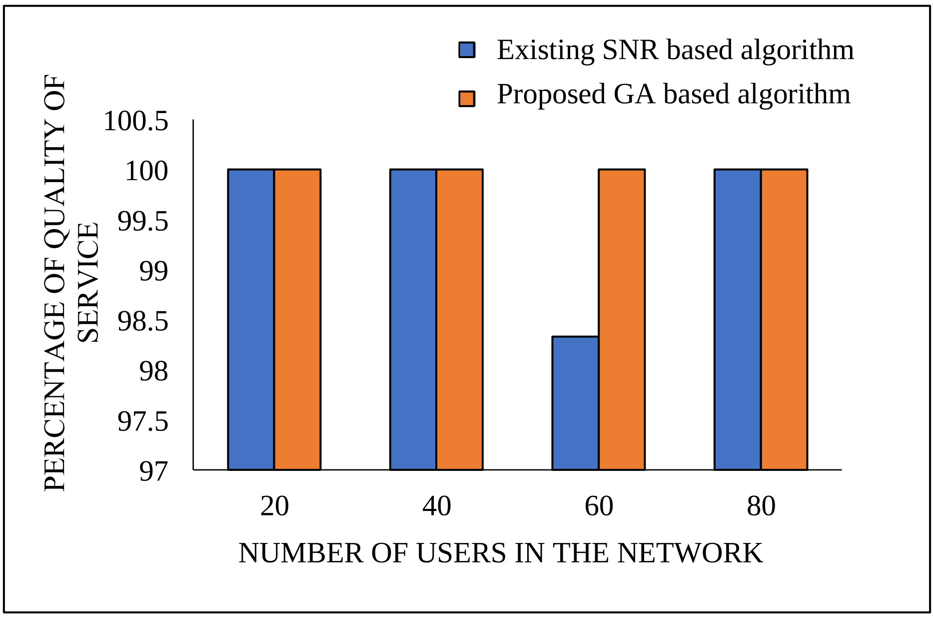



Fig.3.7 Comparison of Quality of Service Between the Existing Algorithm and Proposed Algorithm

In Fig.3.7, the percentage of the ratio between the number of QoS satisfied users associated with the BSs and the total number of user in the HetNet is shown. The proposed scheme is slightly better than the existing SNR based algorithm because the proposed scheme takes into account the QoS requirements of each user. Nonetheless, the performance of the existing SNR based scheme is considered comparable with the proposed scheme, because user association based on highest SNRs can guarantee most of the users with high achievable throughput.

### 3.6.3 Percentage of Users Admitted to the Network

The percentage of the number of users accepted to the network divided by the total number of users *U* is evaluated.

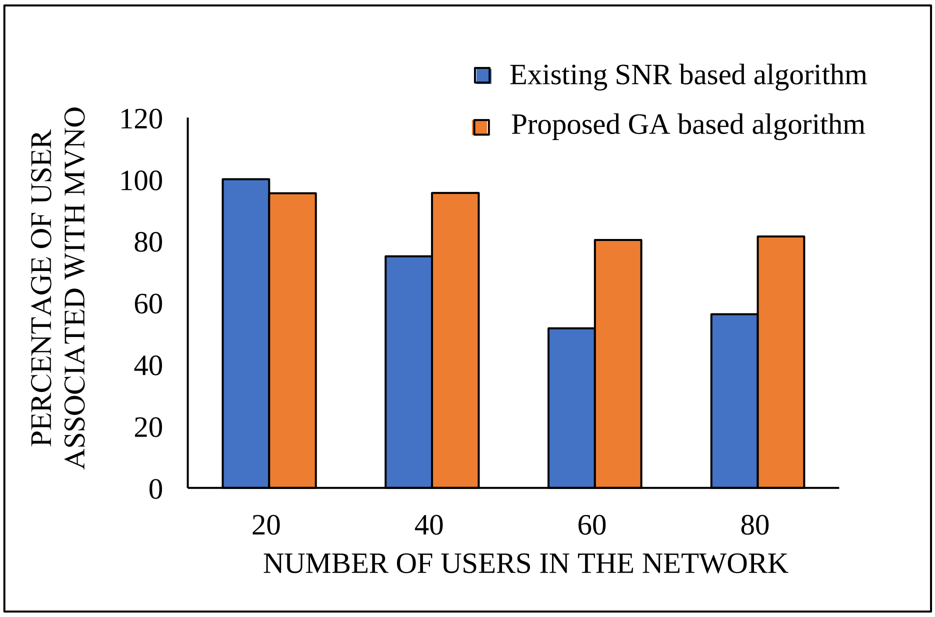



Fig.3.8 Comparison of Percentage of User Associated With MVNO Between the Existing Algorithm and Proposed Algorithm

In Fig. 3.8, the percentage of the number of users associated with the BSs in the Het-Net is shown. The proposed scheme has shown to allow higher numbers of users accepted by the network at all instances except the instance of 20 users. Although, at the instance of 20 users the number of users accepted by the network by the proposed scheme and the existing scheme does not have a vast difference. This is because the existing algorithm is incapable to fulfil the QoS requirements of all users, hence more users are dropped from the network. As a result, the numbers of users admitted to the network for the existing algorithm also fluctuates for all different total numbers of users in the network.

### 3.6.4 Power Consumption

The power consumption is also assessed. At each realization, the power consumption is the sum of power consumption and the transit power of the BS.

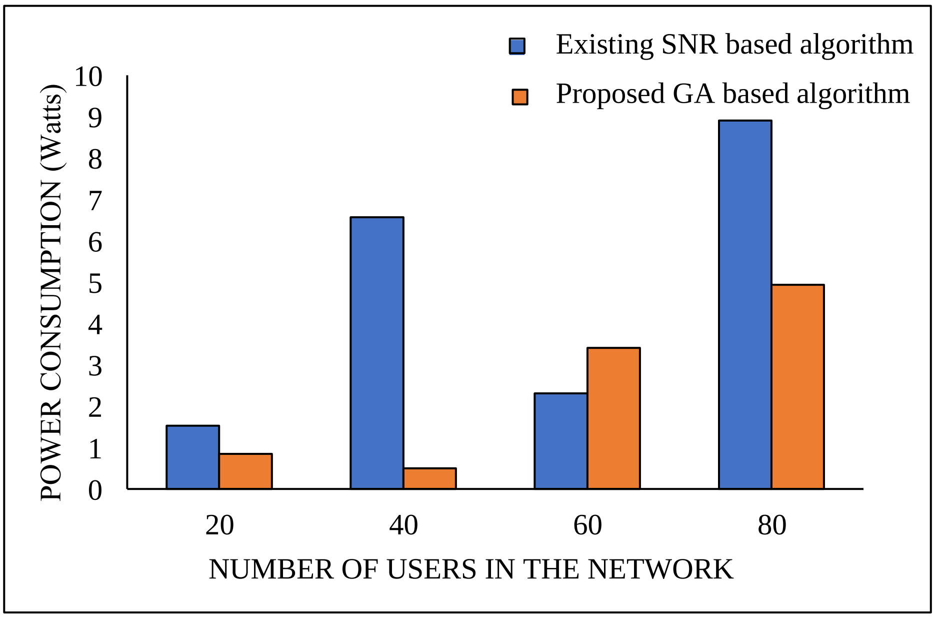



Fig.3.9 Comparison of Power Consumption Between the Existing Algorithm and Proposed Algorithm

In Fig. 3.9, the power consumption by the network is shown. The power consumption at the instance of 40 users is the lowest with the proposed scheme when compared to existing scheme. The proposed scheme is overall the best compared to the existing algorithm because the former aims to optimize energy efficiency. Although, at the instance of 60 users the power consumption with the proposed scheme is bit higher than the existing scheme.

### 3.7 Conclusion

Simulation results showed that the proposed GA-based user association scheme is more efficient than the existing algorithm with the increased fairness index, QoS, percentage of user associated with the MVNO and decreased power consumption (in Watts). It is observed that the proposed GA based user association shows an improvement gain of 14.8% in terms of fairness and additional improvement of 17.5% in terms of percentage of user associated to the network. So, when compared to the existing scheme and the proposed scheme, the latter outperforms than the existing scheme.

# CHAPTER 4

# USER ASSOCIATION FOR NETWORK SLICING IN MULTI-TENANT HETEROGENEOUS NETWORKS WITH MULTIPLE BS PER MVNO

## 4.1 Introduction

As discussed in Chapter 3, the user association problem is formulated by considering the scenario that only one BS is available to a MVNO in one network slice. Furthermore, the previous scenario assumes that the PRBs are orthogonally divided among the BSs, thus intra-slice interference is not available. This chapter considers a scenario in which more than one BSs are associated with a MVNO in a network slice with non-negligible intra-slice interference, which is an extended work of the scenario described in the previous chapter.

## 4.2 System Model

As mentioned in Section 3.2 in chapter 3, there are BSs, MVNOs and users in the Het-Net. The major difference between the user association scenario in previous chapter and this chapter is that there is only one BS is connected to a MVNO in the former scenario, whereas in the current scenario is that there are more than one BSs are associated with a MVNO. In addition to that, $M$ is the set of MVNOs. Similar to the scenario in Chapter 3, radio resources are orthogonally divided among the different MVNOs, and in this chapter, intra slice resource sharing among BSs associated with the same MVNO is considered. The following Fig.4.1 illustrates the system model of the user association in a Het-Net where more than one BS is connected to a MVNO.

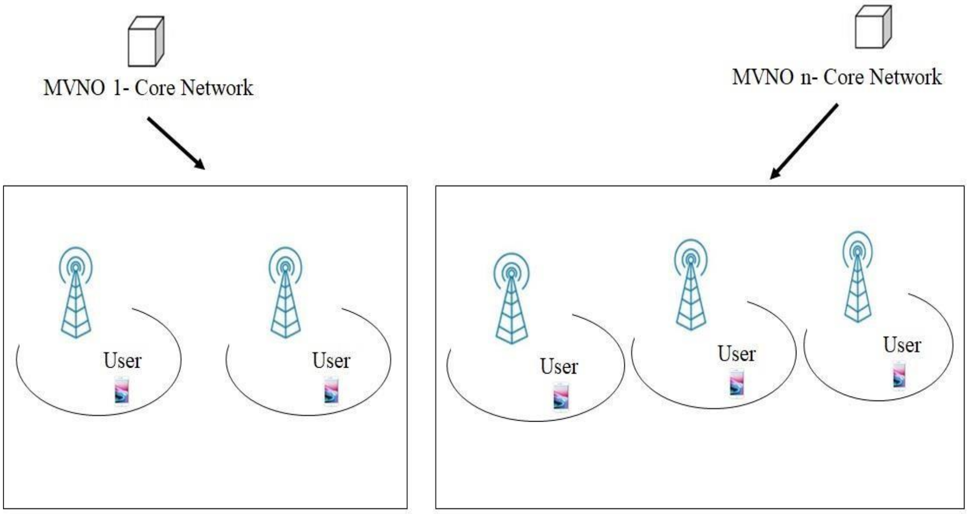



Fig.4.1 Het-Net in which multiple BSs are connected to a MVNO

### 4.3 Problem Formulation

Due to the fact that the BSs associated with the same MVNO can share or access the same radio resources allocated to the MVNO, intra-slice interference, i.e., the interference between the BSs that are associated with the same MVNO needs to be considered.

Let us consider that $a_{ki}$ is the user association variable, where $a_{ki}$ is equal to 1 if the user $i$ connects with BS $k$, 0 otherwise. In the current scenario, more than one BS are associated with each MVNO in one network slice, and the intraslice interference between these BSs occurs when they access the same RBs.

For the sake of capacity modelling simplicity, a worst-case scenario is considered for the estimation of the user capacity, by assuming that each user always experiences interference from all other BSs in the same slice. As such, the worst-case throughput, i.e., minimum achievable throughput of each user *i* on a PRB by BS *k* can be evaluated by

$$R_{ki} = Blog_2(1 + \tau_{ki}) \tag{4.1}$$

Where, *B* is the spectrum bandwidth of a PRB, and

$$\tau_{ki} = \frac{P_{max,k}\, g_{ki}}{I_{ki} + P_{AWGN}} \tag{4.2}$$

is the SINR between the user *i* and the BS *k*. In equation 4.2, the $P_{max,k}$ is the maximum transmission power of BS *k*, $g_{ki}$ is the downlink channel gain between the BS *k* to the user *i* and $P_{AWGN}$ is the additive white Gaussian noise power. It is assumed that the user association is carried out in a time scale which is coarse grained. Furthermore, it is assumed that the channel gain of all PRBs are averaged over the user association period and one user receives only one PRB. $\propto_i$ is the target data rate of each user *i*. Similar to the reason as stated in section 3.4, the data rate is the only QoS requirement considered in this study. $I_{ki}$ is the interference experienced by the user *i* from the BS *k*, which differentiates the scenario explained in the previous chapter from this chapter of the thesis. The equation of $I_{ki}$ is given below:

$$I_{ki} = \sum_{j\in K_m/\{k\}} P_{max,j}\, g_{ji} \tag{4.3}$$

Where, $P_{max,j}$ is the transmit power of the BS $j$, where $j\in K, j \neq k$. It is noteworthy that the actual throughput may be able to be estimated but its estimation is impractical because different PRBs may be allocated to each user during resource scheduling in the lower layer of the network at a fine-grained time scale. Thus, the interference experienced by each user can vary rapidly within the fine-grained time scale. Since user association is usually performed at a coarse-grained time scale, it is reasonable to assume that each user always experiences interference from all other BSs in the same slice and evaluate the user's worst-case throughput.

In this chapter, the user association problem is formulated to maximize the weighted achievable minimum throughput, as follows:

$$max_{(a_{ki})} \sum_{m\in M} \sum_{k\in K_m} \sum_{i\in U} W_k V_i a_{ki} R_{ki} \tag{4.4}$$

subject to

$$\sum_{k\in K} a_{ki} \leq 1, \forall\, i \in U \tag{4.4a}$$

$$\sum_{i\in U} a_{ki} \leq N_k, \; \forall\, k \in K \tag{4.4b}$$

$$a_{ki} \in \{0, 1\} \; \forall\, i \in U, k \in K \tag{4.4c}$$

where $a_{ki}$ is the user association variable, (i.e.,)

$$a_{ki} = \begin{cases} 1 \text{ if user } i \text{ is associated with BS } k \\ 0 \text{ otherwise} \end{cases}$$

and $K_m$ is the set of BSs associated with MVNO $m$. In constraint (4.4a), each user is allowed to associate with only one BS. Constraint (4.4b) ensures that the total number of PRBs must be lesser than or equal to the number of users associated to the BS, as we have assumed that one user only receives one PRB. $W_k$ and $V_{ki}$ are the weightage co-efficient of BS and user, such that $0 \leq W_k \leq 1$; $0 \leq V_{ki} \leq 1$ respectively. The $W_k$ and $V_{ki}$ are set the same as in equation 3.4 and equation 3.5 respectively mentioned in chapter 3.

### 4.4 Proposed Algorithm and Solution

As mentioned in previous chapter, the GA was designed to solve the 0/1 integer programming problem. The equation (4.4) is also 0/1 integer programming problem and hence GA was adopted to solve (4.4). As the Fig.4.1 illustrates that more than one BSs are deployed under each BSs. The GA produces 1, when the BS is connected to a user, and produces 0, when the BS is not connected to a user.

The constraints in (4.4) are converted into a penalty function when GA was adopted to solve it. The constraint in equation (4.4a) and (4.4b) are converted into second and third term of the unconstrained problem in equation

(4.5) given below Hence, the constrained problem will be converted into an unconstrained problem as follows:

$$max_{(a_{ki})} w_1 \sum_{k \in K} \sum_{i \in U} w_k v_{ki}\, a_{ki}\, R_{ki} \;\; +$$

$$w_2 \frac{\sum_{i \in U} \langle 1 - \sum_{k \in K} a_{ki} \rangle}{|U|} + \; w_3 \frac{\sum_{k \in K} \langle N_k - \sum_{i \in U} a_{ki} \rangle}{|K|} \qquad (4.5)$$

Where, $w_1$, $w_2$, $w_3$ are the weighting co-efficient for each term in the objective function, and

$$< z > = \begin{cases} z \; if \; z < 0 \\ 0 \; otherwise \end{cases} \qquad (4.6)$$

The problem in equation (4.6) is solved using GA. It is done similarly as the problem in (3.6) was solved. The GA begins with selection and here, roulette wheel selection is used. Number of chromosomes is denoted by *n*. After the chromosomes are produced, the best chromosomes are chosen and saved for the selection phase, where the parent chromosomes are chosen using a roulette wheel. Two chromosomes from the parent chromosomes are chosen at random for crossover to generate new offspring. Then there is mutation, which involves changing a bit from 0 to 1 or vice versa.

Table 4.1 shows the summary of the proposed GA-based user association scheme.

TABLE 4.1: PROPOSED GENETIC ALGORITHM FOR THE USER ASSOCIATION IN A MULTI-TENANT HETEROGENEOUS NETWORK SLICING.

| | |
|---|---|
| **Step 1** | Generate an initial population matrix 'a' of size (*K***U*), where *K* and *U* are the total number of BS and users in the network respectively. To be noted, the number of BSs and number of MVNOs are the same in the network. |
| **Step 2** | Set the selection rate *sel* = 0.5, mutation rate *mut* = 0.1, number of generations *G* =100, number of chromosomes *n* = 100. Each chromosome is a candidate solution for (4.4). |
| **Step 3** | Compute the fitness function (objective function) using equation (4.5) for each chromosome. |
| **Step 4** | Select the best chromosome (BS) from the initial population according to the selection rate. The genetic algorithm begins with the selection process using Roulette Wheel selection. |
| **Step 5** | The selected chromosomes are the parent chromosomes. |
| **Step 6** | Single point crossover is performed over the selected chromosomes. The crossover point is calculated, the bits from the crossover point to the last bit of the parent chromosomes are swapped between themselves to produce two new offsprings. |

| | |
|---|---|
| | A bit flip mutation is performed over the off-spring chromosomes. The bit which is to be mutated is chosen randomly and that bit is flipped (one to zero or vice versa). |
| **Step 7** | Compute the fitness function (objective function) using equation (4.5) for each chromosome. |
| **Step 8** | Sort out the best fit chromosome by calculating the fitness of chromosomes by comparing the fitness of chromosome of previous generation and the current generation during each generation. |
| **Step 9** | A new generation population (user association) matrix of size (*K**U*) is developed from each chromosome. |
| **Step 10** | Repeat step 4 to step 9 until the number of generations has been reached. |

## 4.5 Results and Discussion

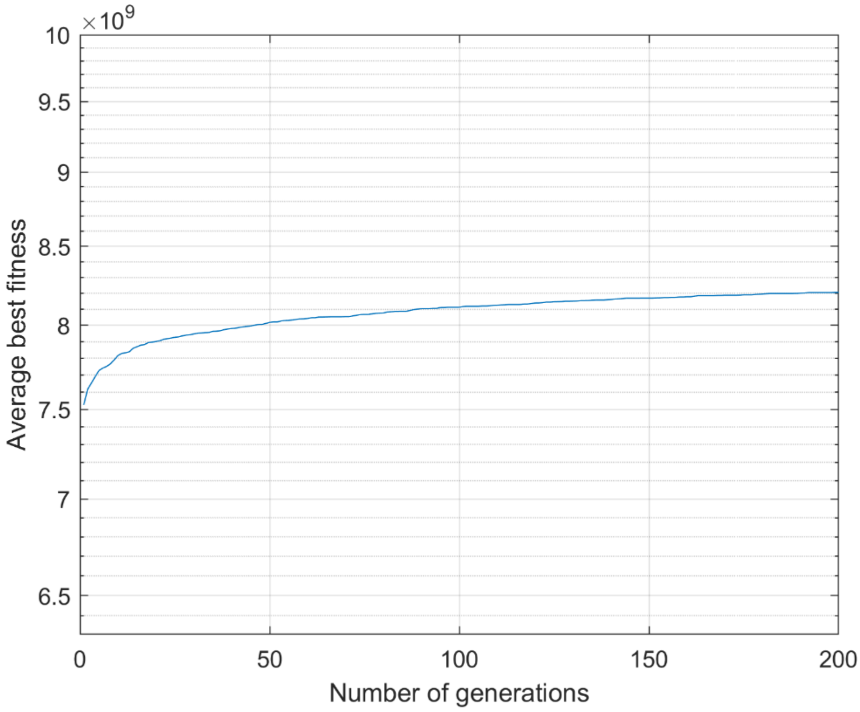
To compare the performance of the performance metrics such as fairness index, quality of service, percentage of user associated with the network, power consumption and energy efficiency, the traditional maximum SINR based user association scheme denoted as 'Existing SINR based algorithm' (Ye *et al*., 2013) was used. Also, the channel model is pathloss channel. The SBS-to-UE pathloss is given as 140.7+36.7log($d$), and MBS-to-UE pathloss is given as 128.1+37.6log ($d$), where $d$ is the distance between user and BS and is given in unit km. The weighting co-efficient values $w_1$, $w_2$ and $w_3$ are obtained experimentally using the trial-and-error method, whereby numerous different values of the weighting co-efficient values are tested and the values that provide the best network performance are selected.

TABLE 4.2: SIMULATION PARAMETERS

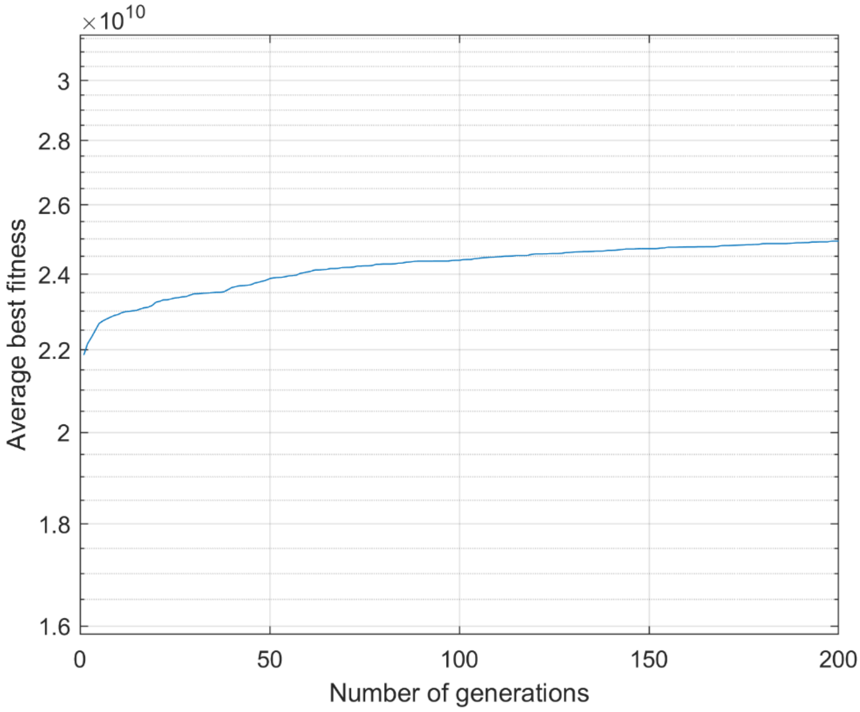
| **Parameters** | **Symbols** | **Values** |
|---|---|---|
| No. of MVNO | $\|M\|$ | 3 |
| No. of BSs per MVNO | $\|K_1\|$, $\|K_2\|$, $\|K_3\|$ | 3, 3, 4 |
| No. of Users | $\|U\|$ | 20,40,60,80 |
| Transmission Power of Base Station | $P_{max}$ | [25, 35] dBm |
| Bandwidth of each PRB | $B$ | 180 kHz |
| Noise figure | $NF$ | 9 dB |
| Noise Spectral Density | $N0$ | -174 dBm |
| Data Rate Requirement | $\propto_i$ | [100, 300] kb/s |
| Selection Rate | $sel$ | 0.5 |

| | | |
|---|---|---|
| Mutation Rate | *mut* | 0.1 |
| No. of chromosomes | *n* | 100 |
| No. of generations | *G* | 100 |
| Weighting Co-efficient | $w_1/w_2/w_3$ | 0.9/0.001/0.099 |

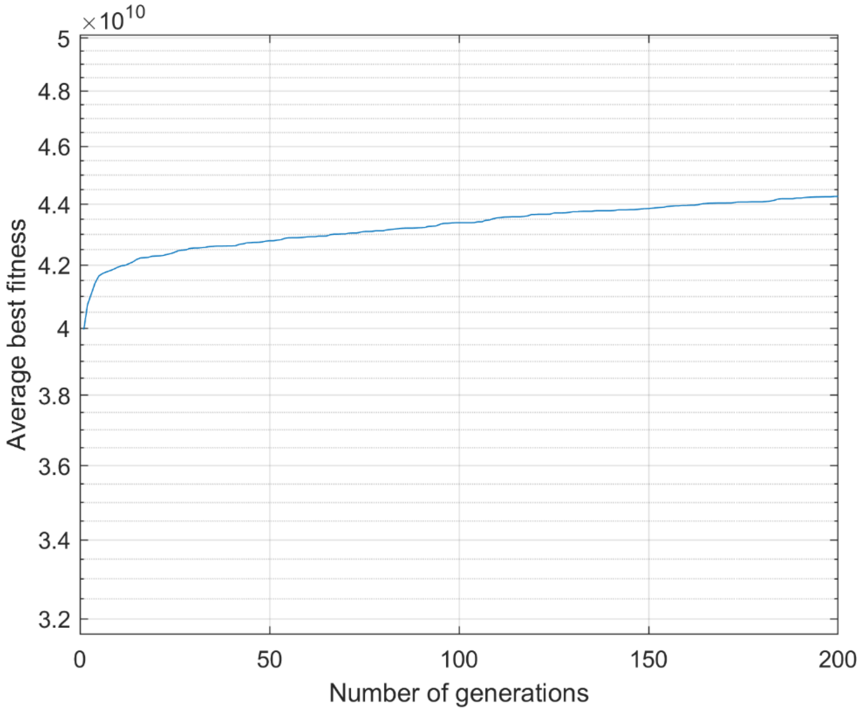
The convergence behaviour of the proposed GA based user association scheme for different number of users (20, 40, 60 and 80) in the network has been illustrated in Fig.4.2 below. It is observed that the convergence of the GA based user association scheme begins at the $50^{th}$ generation on average over 100 runs for all the four scenarios. Then, the proposed scheme is compared with the existing SINR based user association scheme in terms of the aforementioned performance metrics.


×10^9
10
9.5
9
8.5
8
7.5
7
6.5
Average best fitness
0
50
100
150
200
Number of generations

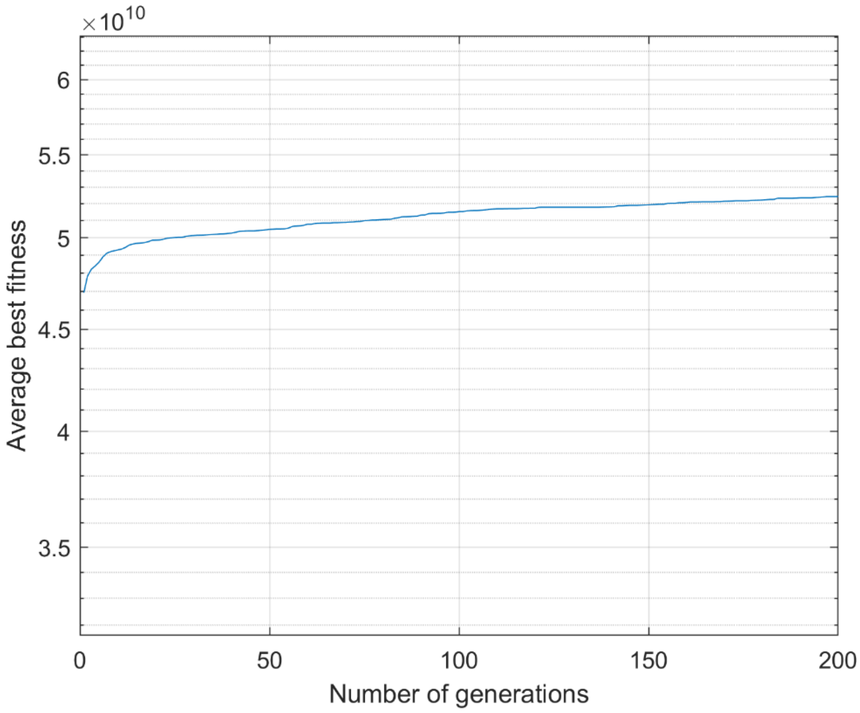
(a)


×10^10
3
2.8
2.6
2.4
2.2
2
1.8
1.6
Average best fitness
0
50
100
150
200
Number of generations

(b)




(c)




(d)

Fig.4.2 Illustration of GA convergence for scenarios with (a) 20, (b) 40, (c) 60 and (d) 80 users.

### 4.5.1 Fairness Index

Jain's fairness index (Lee *et al.*, 2015) is used to calculate the fairness index of the network using the following.

$$Fairness\ Index = \frac{(ATM'1' + \cdots + ATM'n')^2}{|M| * \left(ATM'1'^2 + \cdots + ATM'n'^2\right)} \tag{4.7}$$

Where, *ATM 'n'* is the minimum achievable throughput achieved for MVNO *n*.

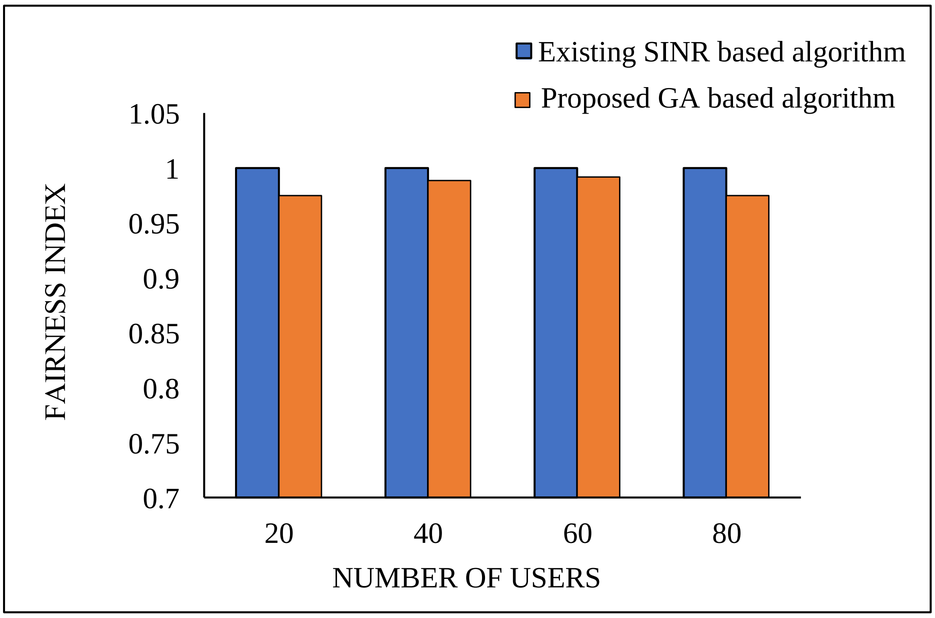



Fig.4.3 Comparison of Fairness Index Between the Existing Algorithm and Proposed Algorithm

Fig. 4.3 depicts the achieved fairness index with various number of users in the system. The proposed scheme is compared with the existing SINR based user association scheme. However, both the schemes performs more or less the same and they are comparable, but the user association performed by the latter completely depends upon the SINR of the users.

### 4.5.2 Quality of Service

Quality of service (QoS) based upon lower bound achievable throughput is evaluated as the percentage of users that achieve the required data rate divided by the total number of users $|U|$ based upon the lower bound or minimum achievable throughput.

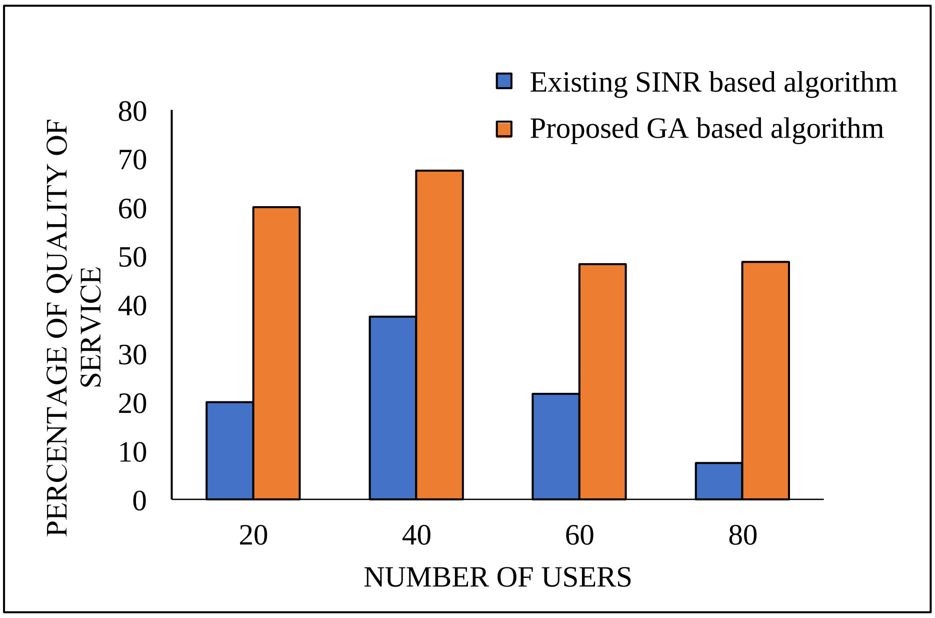



Fig.4.4 Comparison of Quality of Service Between the Existing Algorithm and Proposed Algorithm

Fig. 4.4 shows the comparison of percentage of quality of service with the instances of various number of users in the network. The proposed scheme outperforms the existing scheme because the proposed scheme takes the QoS into consideration. The result shown is based on the minimum or lower bound achievable throughput, which is greater than the target data rate. Thus, the actual throughput and the QoS can be possibly higher than the shown result. It shows that the instance having 40 users has better QoS provisioning when compared to the other instances. Even though the percentage of QoS varies at different instances, the proposed algorithm has the higher QoS percentage when compared to the existing algorithm.

### 4.5.3 Percentage of Users Admitted to the Network

The percentage of the number of users admitted to the network divided by the total number of users |*U*| is evaluated.

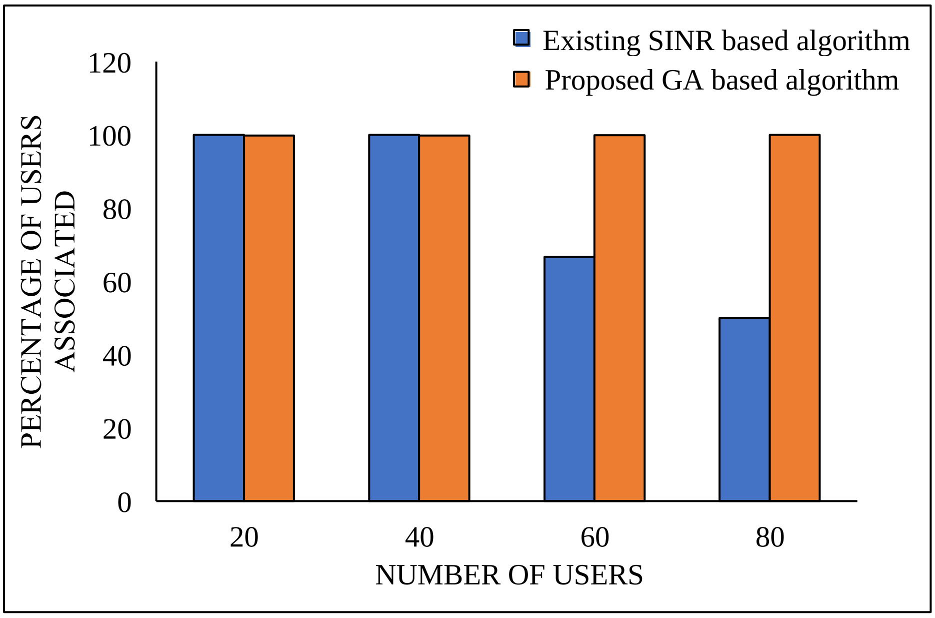



Fig.4.5 Comparison of Percentage of User Associated with MVNO Between the Existing Algorithm and Proposed Algorithm

Fig. 4.5 shows the comparison of percentage of users admitted to the network to the total number of users in the network. The proposed algorithm using GA outperforms the existing SINR based user association scheme. The percentage of users associated with the BS is 100% at every instance with varying number of users. When the number of users in the network increases to 60, the SINR based user association scheme was not capable of accepting more users. This shows that all the users are connected to the BS regardless of the number of the user in the network using the proposed scheme.

### 4.5.4 Power Consumption

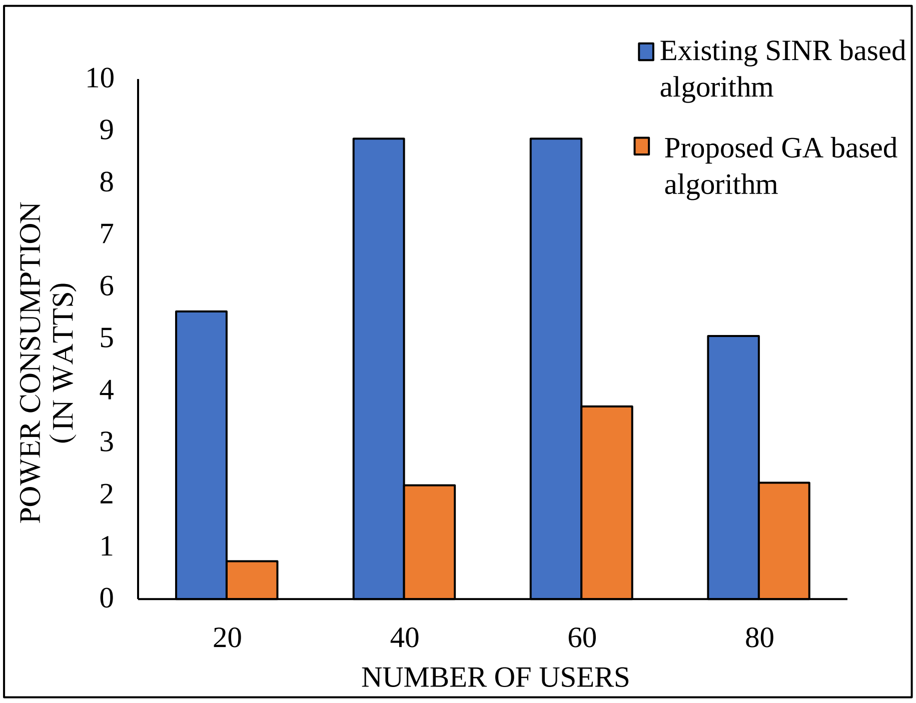



Fig.4.6 Comparison of Power Consumption Between the Existing Algorithm and Proposed Algorithm

Fig.4.6 illustrates the achieved power consumption over various numbers of users. It is shown that the proposed scheme outperforms the existing SINR based user association scheme, as the former aims to minimize the network power consumption. Although, the power consumption when there are 60 users is a bit higher than the other instances, it is still better than the existing scheme. As such, the proposed scheme achieves higher power saving which also leads to a lower energy expenditure.

## 4.6 Conclusion

The simulation results showed that the proposed algorithm and the existing algorithm were compared and proved that the former outperforms in terms of QoS, percentage of users associated with the users and power consumption.  However, in terms of fairness, the proposed scheme was comparable with the existing scheme and there is not much of difference in their performances. It is observed that the proposed GA based user association shows an improvement gain of 68.7% in terms of power consumption and additional improvement of 34.5% in terms of QoS of the network. Thus, the proposed scheme outperforms the existing scheme in the performance metrics forementioned.

# CHAPTER 5

# CONCLUSION AND FUTURE WORK

## 5.1 CONCLUSION

A user association scheme in a multi-tenant Het-Net under two different scenarios such as single BS per MVNO and multiple BS per MVNO has been presented in this thesis.

In the first scenario with single BS per MVNO, the performance of the proposed GA based user association scheme is evaluated and compared with existing SNR based (greedy algorithm) user association scheme. Although, the fairness performance slowly gets lower with the increase in number of users, it is comparable with the performance of existing scheme, as the difference is small. The power consumption of the network at different instances varies. The comparison shows that the proposed scheme outperforms the existing algorithm in terms of QoS, fairness and network power consumption. This indicates that our proposed scheme can guarantee fair throughput distribution among users, while these users can experience reliable connectivity from its associated tenant with low power consumption incurred to the Het-Net. As a result of the low power consumption, the power consumption cost of the network can be reduced.

Similarly, in the second scenario with multiple BSs per MVNO, the proposed GA based user association schemes' performance is evaluated and compared with the existing traditional SINR based user association scheme. Although, there is difference in the fairness performance, it is small and comparable, shows the comparison. The comparison also has proved that the proposed scheme is better in terms of fairness, QoS, percentage of users associated and decreased power consumption. This shows that the users will experience a better throughput the fair allocation of resources with lesser power consumption. As the power consumption is low, the expenditure of the network is low as well.

On this note, this thesis concludes that the overall performance of the proposed GA based user association outperforms the existing user association scheme.

## 5.2 FUTURE WORK

In this thesis, PRBs are allocated to each MVNO and only one PRB is allocated to one user. In future work, allocating more than one PRB to a user can be considered. Allocating more than one PRB to a user helps the user to access applications and cloud data which requires high data rate.

Also, the future work can consider Co-ordinated Multi Point (CoMP). This technology utilizes several BSs simultaneously and data are transmitted simultaneously from multiple BSs to a single user or multiple users. Therefore, the users can achieve a higher data rate.

Besides that, in the future, the user association scheme explained in this thesis, can be further modified as a joint sub-channel allocation and user association scheme. Thus, the scheme will have more parameters which will help to achieve more optimized fairness, QoS, power consumption and also can further maximize the user association.

Moreover, the load balancing concept can be incorporated into the user association scheme explained in this thesis. This can be done by shifting the users from overloaded network slice to under-loaded network slices, to balance the load on each slice of the network, when there are many slices. This can be done to achieve more optimized fairness.

# REFERENCES


Ai, N., Wu, B., Li, B. and Zhao, Z., 2021. 5G heterogeneous network selection and resource allocation optimization based on cuckoo search algorithm. *Computer Communications*, 168, pp.170–177. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0140366421000244.

Albonda, H.D.R. and Perez-Romero, J., 2019. An Efficient RAN Slicing Strategy for a Heterogeneous Network With eMBB and V2X Services. *IEEE Access*, 7, pp.44771–44782. Available at: https://ieeexplore.ieee.org/document/8676226/.

Anon, *Generations of Mobile Networks: Explained* [Online]. Available at: https://justaskthales.com/en/generations-mobile-networks-explained/.

Anon, 2002. Framework of a service level agreement, E.860, *Telecommunication Standardization Sector of International Telecommunication Union*

Bahlke, F., Ramos-Cantor, O.D., Henneberger, S. and Pesavento, M., 2018. Optimized Cell Planning for Network Slicing in Heterogeneous Wireless Communication Networks. *IEEE Communications Letters*, 22(8), pp.1676–1679. Available at: https://ieeexplore.ieee.org/document/8368293/.

Buzzi, S. et al., 2016. A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. *IEEE Journal on Selected Areas in Communications*, 34(4), pp.697–709. Available at: http://ieeexplore.ieee.org/document/7446253/.

Caballero, P., Banchs, A., de Veciana, G. and Costa-Perez, X., 2017. MultiTenant Radio Access Network Slicing: Statistical Multiplexing of Spatial Loads. *IEEE/ACM Transactions on Networking*, 25(5), pp.3044–3058. Available at: http://ieeexplore.ieee.org/document/7982693/.

Carr, J., 2014. *An Introduction to Genetic Algorithms*, WHITMAN COLLEGE.

Dai, B. and Yu, W., 2016. Energy Efficiency of Downlink Transmission Strategies for Cloud Radio Access Networks. *IEEE Journal on Selected Areas in Communications*, 34(4), pp.1037–1050. Available at: http://ieeexplore.ieee.org/document/7437385/.

Elayoubi, S.E., Jemaa, S. Ben, Altman, Z. and Galindo-Serrano, A., 2019. 5G RAN Slicing for Verticals: Enablers and Challenges. *IEEE Communications Magazine*, 57(1), pp.28–34. Available at: https://ieeexplore.ieee.org/document/8613270/.

Ge, X., Yang, J., Gharavi, H. and Sun, Y., 2017. Energy Efficiency Challenges of 5G Small Cell Networks. *IEEE Communications Magazine*, 55(5), pp.184–191. Available at: http://ieeexplore.ieee.org/document/7926940/.

Guo, T. and Suarez, A., 2019. Enabling 5G RAN Slicing With EDF Slice Scheduling. *IEEE Transactions on Vehicular Technology*, 68(3), pp.2865–2877. Available at: https://ieeexplore.ieee.org/document/8624268/.

Habibi, M.A., Han, B., Nasimi, M. and Schotten, H.D., 2018. The Structure of Service Level Agreement of Slice-based 5G Network. *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. 2018 IEEE, Bologna, Italy.

Kanwal, K., Safdar, G.A., Ur-Rehman, M. and Yang, X., 2017. Energy Management in LTE Networks. *IEEE Access*, 5, pp.4264–4284. Available at: http://ieeexplore.ieee.org/document/7888442/.

Keith, R., 2019, *5G Key Technologies* [Online]. Available at: https://www.a10networks.com/blog/5g-key-technologies/.

Khalili, A., Akhlaghi, S., Tabassum, H. and Ng, D.W.K., 2020. Joint User Association and Resource Allocation in the Uplink of Heterogeneous Networks. *IEEE Wireless Communications Letters*, 9(6), pp.804–808.

LeAnh, T., Tran, N.H., Ngo, D.T. and Hong, C.S., 2017. Resource Allocation for Virtualized Wireless Networks with Backhaul Constraints. *IEEE Communications Letters*, 21(1), pp.148–151. Available at:

http://ieeexplore.ieee.org/document/7590070/.

Lee, Y.L., Loo, J. and Chuah, T.C., 2015. Modeling and performance evaluation of resource allocation for LTE femtocell networks. In: *Modeling and Simulation of Computer Networks and Systems*. Elsevier, pp. 683–716.

Liberg, O. and Wikström, G., 2020. Physical Resource Block. In: *Cellular Internet of Things*. Science Direct.

Lipowski, A. and Lipowska, D., 2012. Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications*, 391(6), pp.2193–2196. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0378437111009010.

Ma, L. et al., 2018. A Biological Model for Resource Allocation and User Dynamics in Virtualized HetNet. *Wireless Communications and Mobile Computing*, 2018, pp.1–11. Available at: https://www.hindawi.com/journals/wcmc/2018/1745904/.

Matthisen, B., Aydin, O. and Jorswieck, E.A., 2018. Throughput and EnergyEfficient Network Slicing. *WSA 2018; 22nd International ITG Workshop on Smart Antennas*. 2018 VDE, Bochum, Germany, pp. 1–6.

Nie, G., Tian, H. and Ren, J., 2015. Energy Efficient Forward and Backhaul Link Optimization in OFDMA Small Cell Networks. *IEEE Communications Letters*, 19(11), pp.1989–1992. Available at: http://ieeexplore.ieee.org/document/7222400/.

Nie, G., Tian, H., Sengul, C. and Zhang, P., 2017. Forward and Backhaul Link Optimization for Energy Efficient OFDMA Small Cell Networks. *IEEE Transactions on Wireless Communications*, 16(2), pp.1080–1093. Available at: http://ieeexplore.ieee.org/document/7776881/.

Oladejo, S.O. and Falowo, O.E., 2017. 5G network slicing: A multi-tenancy scenario. *2017 Global Wireless Summit (GWS)*. October 2017 IEEE, pp. 88–92.

Ordonez-Lucena, J. et al., 2017. Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges. *IEEE Communications Magazine*, 55(5), pp.80–87. Available at: http://ieeexplore.ieee.org/document/7926921/.

Ramantas, K. et al., 2018. Implementation of an SDN-enabled 5G experimental platform for core and radio access network support. *Advances in Intelligent Systems and Computing*, 725(February), pp.791–796.

Rizvi, S. et al., 2017. An investigation of energy efficiency in 5G wireless networks. *2017 International Conference on Circuits, System and Simulation (ICCSS)*. July 2017 IEEE, pp. 142–145.

Samdanis, K., Costa-Perez, X. and Sciancalepore, V., 2016. From network sharing to multi-tenancy: The 5G network slice broker. *IEEE Communications Magazine*, 54(7), pp.32–39. Available at: http://ieeexplore.ieee.org/document/7514161/.

Sharma, S. et al., 2021. Key Enabling Technologies of 5G Wireless Mobile Communication. *Journal of Physics: Conference Series*, 1817(1), p.012003. Available at: https://iopscience.iop.org/article/10.1088/1742-6596/1817/1/012003.

Shi, S., Yang, W., Zhang, J. and Chang, Z., 2015. Review of Key Technologies of 5G Wireless Communication System Li, J.Y., Liu, T.Y., Deng, T. and Tian, M., (eds.). *MATEC Web of Conferences*, 22, p.01005. Available at: http://www.matec-conferences.org/10.1051/matecconf/20152201005.

Tran, T.D. and Le, L.B., 2020. Resource Allocation for Multi-Tenant Network Slicing: A Multi-Leader Multi-Follower Stackelberg Game Approach. *IEEE Transactions on Vehicular Technology*, 69(8), pp.8886–8899. Available at: https://ieeexplore.ieee.org/document/9099033/.

Vo, P.L., Nguyen, M.N.H., Le, T.A. and Tran, N.H., 2018. Slicing the Edge: Resource Allocation for RAN Network Slicing. *IEEE Wireless Communications Letters*, 7(6), pp.970–973. Available at: https://ieeexplore.ieee.org/document/8370043/.

Wang, Q. et al., 2018. Energy-Efficient Priority-Based Scheduling for Wireless Network Slicing. *2018 IEEE Global Communications Conference (GLOBECOM)*. December 2018 IEEE, pp. 1–6.

Wu, J. et al., 2015. A Study on the Dependability of Software Defined Networks. *Proceedings of the 2015 International Conference on Materials Engineering and Information Technology Applications*. 2015 Atlantis Press, Paris, France.

Xu, Y., Gui, G., Gacanin, H. and Adachi, F., 2021. A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges. *IEEE Communications Surveys and Tutorials*, 23(2), pp.668–695.

Ye, Q. et al., 2013. User Association for Load Balancing in Heterogeneous Cellular Networks. *IEEE Transactions on Wireless Communications*, 12(6), pp.2706–2716. Available at: http://ieeexplore.ieee.org/document/6497017/.

Yuxiang, Z., 2015. *NFV: Making Networks More Agile*. [online] Available at: https://www.zte.com.cn/global/about/magazine/ztetechnologies/2015/2/en_688/432510.html

Zhao, H., Zhao, L., Liang, K. and Pan, C., 2017. Radio access network slicing based on C/U plane separation. *China Communications*, 14(12), pp.134–141. Available at: http://ieeexplore.ieee.org/document/8246330/.

Zhang, W., Hardjawana, W. and Vucetic, B., 2017. User-Base Stations Association in Multi-Tenant Base Station Networks. *GLOBECOM 2017 – 2017 IEEE Global Communications Conference*. December 2017 IEEE, pp. 1–6.

**PUBLICATIONS**

Jayanthi, S.S., Lee, Y.L. and Chang, Y.C., 2021. User Association for MultiTenant Heterogeneous Network Slicing Using Genetic Algorithm. *2021 8th International Conference on Computer and Communication Engineering (ICCCE)*. 22 June 2021 IEEE, pp. 326–330.

Swathi Subbiah Jayanthi, Lee Ying Loong, Yoong Chang Choon, Teong Chee Chuah, Jonathan Loo, Yin Ping Yeck, and Rudzidatul Akman Dziyauddin., 2021. Evolutionary Optimization-Based User Association for Heterogeneous Networks with Multi-tenant Network Slicing. *KSII Transactions on Internet and Information Systems* (Status – Under Review).

*Note: Data and some figures in the above papers have been repeated in this thesis as the publications are from the same research described in this thesis and both papers are written by the author of this thesis