

Mobility Management in 5G Heterogeneous Networks: A Handover Scheme for Reducing Handover Failures



Presented by

Reitumetse Patrice Monaheng

Supervisor:

Dr Daniel Ramotsoela, University of Cape Town (UCT)

Co-Supervisor:

Dr Albert Lysko, Council for Scientific and Industrial
Research (CSIR)

This dissertation is presented to the Faculty of Engineering and
The Built Environment at the University of Cape Town in partial
fulfilment of the academic requirements for the degree of
**Master of Engineering (M.Eng.) in
Telecommunications.**

9th December 2022

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

1. I know that plagiarism is a serious form of academic dishonesty.
2. I have read the document about avoiding plagiarism, am familiar with its contents and have avoided all forms of plagiarism mentioned there.
3. Where I have used the words of others, I have indicated this by the use of quotation marks.
4. I have referenced all quotations and properly acknowledged other ideas borrowed from others.
5. I have not and shall not allow others to plagiarise my work.
6. I declare that this is my own work.
7. I am attaching the summary of the Turnitin match overview (when required to do so).

Signature:

Signed by candidate

Date: 9 December. 2022

I dedicate this to my dearest mother.

Abstract

Mobile/cellular communications have become very popular and advanced significantly in recent decades. Communications will inevitably evolve into the next generation of wireless communications, in which users will be connected via heterogeneous networks. Small cell (SC)-based ultra-dense heterogeneous networks (HetNets), which are underlaid on the coverage of a macro cell, are among the most promising alternatives for increasing capacity and coverage in 5G cellular networks. An ultra-dense network (UDN) refers to a setup in which the density of Radio Access Technologies (RATs) in a geographical area is increased. As a result, the areas covered by individual RATs begin to overlap.

UDNs are regarded as a critical technology for 5G due to their capability to enhance connection quality and expand system capacity. There, small base stations (SBS) are located close to each other in a UDN. As a result, signals from two or more SBS can be received by a single user equipment (UE). This could lead to severe inter-cell interference. This usually happens when a handover has been delayed. If that happens, then, the handover command message (HCM) will not be received by the UE from its serving BS and, a handover failure (HOF) will be declared. Inter-cell interference is so severe in dense small cells that it occurs frequently, thus degrading signal quality and hence resulting in poor services to users.

This dissertation focuses on reducing handover failures due to the unavailability of resources at the target cell when a user equipment moves out of small cells, which have the highest rate of handover failure. By utilising a semi-Markov mobility prediction algorithm for handover management, we have implemented a handover scheme that reduces the number of handover calls dropping. This paves the way for resources in the target base station to be reserved beforehand and thus reducing the number of handover failures significantly.

The results of the proposed scheme were validated with a simulation in MATLAB with an environment consisting of small cells with different radio access technologies. From the simulation results, the prediction of the next location of the user equipment yielded lower handover call dropping as compared to new calls blocking as they were not predicted and hence resources not reserved for them.

Acknowledgements

So long and thanks for all the fish.

Douglas Adams

Table of Contents

Declaration	ii
Abstract	iv
Acknowledgements	v
Table of Contents.....	vi
List of Figures	viii
List of Tables	ix
List of Abbreviations	x
List of Symbols	xii
1 Introduction	1
1.1 Overview.....	2
1.1.1 Mobility Management in 5G Networks	3
1.2 Problem Statement.....	4
1.3 Research Objectives.....	5
1.4 Research Questions.....	5
1.5 Scope and Limitations	5
1.5.1 Scope	5
1.5.2 Limitations	6
1.6 Dissertation Outline	6
2 Literature Review	7
2.1 Background.....	7
2.1.1 Mobility Management in Wireless Networks	8
2.1.2 Mobility Management in 4G	10
2.1.3 Towards 5G Mobile Network	11
2.1.4 Handover Management in 5G	12
2.1.4.1 5G Baseline Handover.....	13
2.1.4.2 5G Conditional Handover (CHO).....	16
2.2 Related Works.....	17
2.2.1 SDN and /NFV – Based Solutions	17
2.2.2 Mobility Management Solutions in Sliced Environments	19
2.2.3 State-Of-The-Art (SOTA) in 5G Mobility Management Solutions	20
2.3 Chapter summary	24
3 5G Heterogeneous Networks Mobility Management Scheme	25
3.1 Heterogeneous Networks and Small cells	25
3.1.1.1 Femto Cell	25

3.1.1.2	Pico Cell	26
3.1.1.3	Micro Cell.....	26
3.1.1.4	Macro Cell	26
3.2	Analytical System Model.....	27
3.2.1	Network Model	27
3.2.2	Propagation and Channel Model	27
3.2.3	Traffic and Mobility Model.....	28
3.3	Problem Formulation	29
3.4	Mobility Prediction	29
3.4.1	Mobility Prediction schemes	29
3.4.1.1	Markov Chain	30
3.4.1.2	Hidden Markov Model	31
3.4.1.3	Artificial Neural Networks	31
3.4.1.4	Bayesian Network.....	31
3.4.1.5	Data Mining	31
3.4.2	Proposed Mobility Prediction Scheme	32
3.5	Admission Control Algorithm	32
3.6	Base Station Selection Algorithm.....	33
3.7	Chapter Summary	38
4	Simulation and Results.....	39
4.1	The Simulation Environment.....	39
4.2	Evaluation Metrics	40
4.3	Simulation Results and Evaluation	42
4.3.1	Result of changing the call arrival rate on the new call blocking probability and handover call dropping probability	42
4.3.2	Result of increasing network capacity on the new call blocking probability and handover call dropping probability	43
4.3.3	Result of increasing departure rate of handover calls on the new call blocking probability and handover call dropping probability.	44
4.3.4	Result of changing the basic bandwidth unit of video calls on the new call blocking probability and handover call dropping probability.	45
4.3.5	Effect of including mobility prediction versus a system with no mobility prediction	47
4.4	Chapter Summary	47
5	Conclusion and Recommendations	48
5.1	Conclusion	48
5.2	Recommendations.....	49
	References.....	50

List of Figures

Figure 1.1 Evolution of Mobile Networks: From 1G to XG	1
Figure 2.1 5G Usage Scenarios and Key Capabilities [57].....	8
Figure 2.2 Classifications of Global Wireless Technologies [43].....	9
Figure 2.3 LTE Handover [61]	10
Figure 2.4 UE state machine and state transitions between NR/5GC, E-UTRA/EPC and E-UTRA/5GC.....	13
Figure 2.5 The Baseline HO Procedure in 5G [31]	14
Figure 2.6 5G NR Mobility Architecture and HO Use Cases [19].....	16
Figure 2.7 The 3GPP 5G Conditional Handover Procedure [31].....	17
Figure 2.8 Classification of SDN and NFV Mobility Solutions.....	19
Figure 3.1 System Architecture of Intra-system and Inter-system Roaming in Different Types of HetNets	25
Figure 3.2 Heterogenous 5G Network System Model [44].....	27
Figure 3.3 System Propagation Model	28
Figure 3.4 Proposed Handover Management Scheme.....	29
Figure 3.5 An Overall Structure of Mobility Prediction [60].....	30
Figure 3.6 Proposed Handover Scheme	33
Figure 3.7 Incoming, Terminated and Outgoing Calls in a Cell	34
Figure 3.8 Base Station Selection and Call Admission Control Algorithm	37
Figure 4.1 Handover from RAT1 to RAT2	41
Figure 4.2 Impact of changing the call arrival rate on the call blocking probability and handover call dropping probability.....	43
Figure 4.3 Impact of increasing network capacity on the new call blocking probability and handover call dropping probability.....	44
Figure 4.4 Impact of increasing departure rate on the new call blocking probability and handover call dropping probability.....	45
Figure 4.5 Impact of changing the basic bandwidth unit of video calls the new call blocking probability and handover call dropping probability.	46

List of Tables

Table 1.1 5G Handover Management Schemes Comparisons	4
Table 2.1 Classifications of Global Wireless Technologies.....	9
Table 2.2 Comparison between the Mobility Requirements for 4G and 5G [39]	12
Table 2.3 Handover Measurement Report Triggering Events.....	13
Table 2.4 Overview and Qualitative Evaluation of Handover Schemes	22
Table 3.1 Comparison of Small Cells.....	26
Table 4.1 Simulation Parameters.....	39

List of Abbreviations

1G	First Generation of Mobile Networks
2G	Second Generation of Mobile Networks
3G	Third Generation of Mobile Networks
3GPP	Third Generation Partnership Project
4G	Fourth Generation of Mobile Networks
5G	Fifth Generation of Mobile Networks
6G	Sixth Generation of Mobile Networks
AI	Artificial Intelligence
AP	Access Point
ANN	Artificial Neural Network
AR	Augmented Reality
BBU	Baseband Unit
BS	Base Station
CA	Carrier Aggregation
CHO	Conditional Handover
CP	Control Plane
DC	Dual Connectivity
DSL	Digital Subscriber Line
DMM	Distributed Mobility Management
eMBB	enhanced Mobile Broadband
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FTP	File Transfer Protocol
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HCM	Handover Command Message
HetNet	Heterogeneous Network
HO	Handover
HOF	Handover Failure
HOM	Handover Margin
HOPP	Handover Ping-Pong
IoT	Internet of Things
IP	Internet Protocol
ISDN	Integrated Services Digital Network
LA	Location Area
LAN	Local Area Network
LOS	Line of Sight
LTE	Long-Term Evolution
MAB	Multi-Armed Bandit
MAN	Metropolitan Area Network
MCC	Mobile Country Code

MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
MME	Mobility Management Entity
mMTC	massive Machine-Type Communications
MNC	Mobile Network Code
MRO	Mobility Robustness Optimization
MSC	Mobile Switching Center
NG	Next Generation
NR	New Radio
NSA	Non- Stand Alone
NFV	Network Functions Virtualization
PAN	Personal Area Network
PDU	Protocol Data Unit
PLMN	Public Land Mobile Network
PPP	Poisson Point Process
PSTN	Public Switched Telephone Network
QoE	Quality of Experience
QoS	Quality of Service
RA	Registration Area
RAN	Radio Access Network
RAT	Radio Access Technology
RLF	Radio Link Failure
RRC	Radio Resource Control
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RSRP	Received Signal Reference Power
SA	Stand Alone
SC	Small Cell
SDN	Software Defined Networking
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
S-WG	Serving Gateway
TA	Tracking Area
TTT	Time to Trigger
UDN	Ultra-Dense Network
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UP	User Plane
URLLC	Ultra-Reliable, Low-Latency Communications
VHD	Vertical Handover Decision
VLC	Visible Light Communication
VR	Virtual Reality
WAN	Wide Area Network
WLAN	Wireless Local Area Network

List of Symbols

Symbol	Description	Value/Units
P	Power	Decibel-milliwatts (dBm)
μ	Departure Rate	Call per Second (Calls/Minute)
ρ	Cell Load	Percentage (%) usage of total resources
G	Gain	Decibels (dB)
λ	Arrival Rate	Call per Second (Calls/Minute)
σ	Shadow Fading Margin	Decibels (dB)

Chapter 1

Introduction

Radio technologies have demonstrated tremendous development since the introduction of the first-generation (1G) analogue cellular network about forty years ago. Subsequently, communication networks are steadily undergoing a transformation approximately every tenth year to realise the increasing data demands: 2G (the 1990s), 3G (2000s), 4G (2010s), 5G (2020), and 6G is expected to be commercially launched in 2030. The second-generation ushered in the digital era - where voice calls were able to be encrypted, enabled effective use of radio spectrum, and data transmission upgraded to match dialup and Digital Subscriber Line (DSL) offerings. 3G offered digital communication networks better bandwidth as evidenced by the proliferation of smartphones and similar devices. The fourth generation ushered in an all-IP based technology with data speeds of up to 1 Gbps [1]. The all-new 5G technology, which is envisioned to provide ubiquitous connectivity, is expected not solely to work as an enhancement of cellular broadband networks but correspondingly to introduce exclusive network and service capabilities. To begin with, it will ensure a high level of experience in difficult circumstances such as high mobility (like in high-speed trains), densely or sporadically populous areas, and areas encompassed by heterogeneous access technologies. Additionally, 5G is also touted to be an enabler of the Internet of Things (IoT) by provisioning sensors and actuators with access to low energy and low power platforms. Moreover, services requiring real-time and high reliability with low latency will rely heavily on 5G infrastructure [2]. Figure 1.1 below shows how mobile networks have evolved with corresponding generations and capabilities.

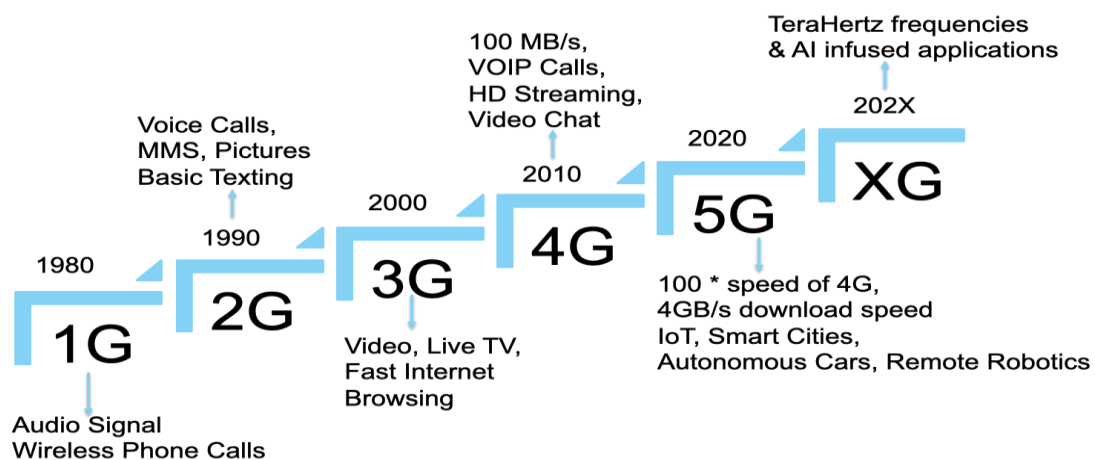


Figure 1.1 Evolution of Mobile Networks: From 1G to XG

Because of the present global pandemic, the amount of data sent over mobile devices is expected to skyrocket in the coming years, which has increased digital adoption the world over, amplified smartphone usage, tablets and numerous gadgets operating a substantial variety of video and conference applications. According to Ericsson Mobility Report [14], there are over 6 billion unique mobile subscribers globally, and mobile subscriptions grew around 3 percent year-on-year, even though there was a slight decrease during the second quarter of 2020 due to the prevailing global pandemic; 80 million of those subscriptions are with 5G-capable devices.

It has been stated that obtaining a new spectrum from telecommunications regulators is an extremely expensive and time-consuming solution to achieving such a massive mobile traffic load [49][50]. Hence, telecommunication operators are moving towards substitute solutions like densified small cells to accomplish enormous growth in wireless network volume and to capitalise on frequency reuse. However, the downside to small cell deployment is the increase in the number of handovers since small cells cover a small geographical area. The subsequent sections will give a brief overview and a summary of mobility management and more specifically 5G mobility management.

1.1 Overview

Mobility Management is a set of technologies that allow continuous access to wireless systems and mobile services while the user is in motion without an interruption of such services. Mobility in wireless systems can be categorised as follows [6]:

- *Terminal Mobility*: In wireless access networks, it refers to the capability of a user device, while in motion, to access and use telecommunications services.
- *User Mobility*: Describes the ability of a mobile device to have access to telecommunication services from diverse base stations under a similar identity when in motion.
- *Service Mobility*: Relates to the capability that a user terminal can be able to access identical telecommunication services irrespective of its location.

According to [7], fundamental prerequisites for future mobility management schemes incorporate support for all categories of mobility; seamless mobility across various access networks with no interruptions in service, time-sensitive applications, uninterrupted ongoing sessions across different network domains as well as global roaming. To achieve the above-mentioned requirements, next-generation (NG) wireless networks should have three basic functional capabilities for Mobility Management [8]:

- *Location Management*: dealing with Location planning, Paging, Addressing, Location tracking and updating.
- *Handover Management*: catering for Signal strength variation, Interference, Call dropping rate, Handover initiation, Connection routing, and Handover algorithms.
- *Resource Management*: addressing Admission Control, Allocation schemes, blocking incoming calls, Per-cell traffic flow, Quality of Service (QoS), Signalling and Estimation of load.

Mobility can also be characterised as either radio mobility (Terminal/ User Mobility) or network mobility (Service Mobility), where radio mobility is primarily dealing with handover management, while network mobility mostly deals with location management (updating and paging). As mobile terminals travel across network boundaries, whether it is a cell, site, or a mobile switching centre (MSC) area, the system must have the ability to

discover the mobile terminal, allocate enough resources to provide seamless continuing service and automatically route the call to the user/terminal. Radiocommunication networks offer services to subscribers by employing the use of base stations, which are accountable for transmitting communication services to the mobile nodes in its cell. A cluster of cells forms a Location Area (LA)/Routing Area/Registration Area (RA)/Tracking Area (TA), managed by a Mobile Switching Centre (MSC), which is linked to the Public Switched Telephone Network (PSTN)/Public Land Mobile Network (PLMN). PLMN is the mobile equivalent of the PSTN, hence it is a system of all mobile wireless networks that use earth-based stations rather than satellites. Such systems can be standalone, or they can be interconnected to fixed networks like the PSTN. PLMNs are the most extensively used networks in the world; examples are the Global System for Mobile Communications (GSM), General Packet Radio Service (GPRS), and the Universal Mobile Telecommunications System (UMTS) also known as 3G, 4G Long-Term Evolution (LTE) and 5G New Radio (NR). An individual PLMN is defined as a region that has a particular Mobile Country Code (MCC) and Mobile Network Code (MNC) it is linked to.

1.1.1 Mobility Management in 5G Networks

Fifth generation (5G) Mobility Management takes the form of intra-domain and inter-domain mobility management structures which describe the movement of a mobile node (MN) within the same domains and between different domains, respectively. The 5G mobility scheme is like the intra-system and inter-system handover process in 4G. However, the intra-domain will focus mainly on fast, reliable, seamless mobility support within a given area of coverage. The inter-domain, on its part, is a scheme that provides a universal mobility solution with rewards of reliability, flexibility, mobility robustness, and scalability. The scheme should be flexible enough for users to join and leave any network without experiencing any service interruption and so enjoy transparent mobility support [15].

Mobility is undeniably a fundamental aspect of wireless communication networks. Seamless mobility spanning several cell territories is attained by a handover scheme in cellular systems. In the past decade, much research has focused on reducing unnecessary handovers, handover ping pong and reducing signalling overheads when executing a handover. The authors in [24] have proposed a method for enabling handovers with slice assistance in multi-RAT networks that include 3GPP NR and non-3GPP WLAN by extending the multi-connectivity protocol. In [25], a Software Defined Network (SDN)-based Quality of Service (QoS) Vertical Handover Decision (VHD) where the network with the highest Received Signal Strength (RSS) and highest QoS score is selected for handover (HO) in dense Heterogeneous Networks (HetNets). They argue that the SDN paradigm is envisioned to improve the prospective performance of heterogeneous networks, particularly the mobility management of users moving through densely packed networks. SDN reduces signalling overhead and HO latency without complicating existing network nodes. The author in [28] focuses on reducing handover latency in 5G mmWave networks.

The work in [46] proposed a handover scheme that reduces handover failure by minimising interruption time for handover preparation and using a threshold margin parameter as a handover execution criterion. Table 1.1 below shows the summary of selected mobility management schemes for 5G heterogeneous networks.

Table 1.1 5G Handover Management Schemes Comparisons

Paper	Environment	Comments/Gaps
[24]	Network Slicing, multi-RAT (Radio Access Technology) networks	The work is aimed at 5G multi-RATs, but the simulation is carried out for an LTE network and an assumption is made that it will also work for 5G deployment.
[25]	Software-Defined Networking (SDN), Heterogeneous Network (HetNet)	The authors proposed an algorithm for a vertical handover from a Macro Cell to a Femto Cell but there is no provision for handling handover failure
[26]	HetNet	The authors designed a velocity-based algorithm to improve handover performance under different mobility scenarios. The solution does not consider a load balancing mechanism for both the serving and target base stations.
[27]	HetNet	The authors proposed a spatiotemporal mobility prediction scheme with a load balancing mechanism for HetNets, but no mechanism for handling handover failures.
[29]	Vehicle to Everything (V2X) network architecture leveraging SDN and Network Slicing using bargaining games	The primary goal of the work is to reduce handover failure. A resource borrowing scheme is designed for inter-slice handover and modelled using bargain games. Unnecessary handovers are not catered for in case of the vehicle running at high speed and passing Roadside Units (RSUs) quickly
[30]	mmWave, HetNet, leveraging Multi-Armed Bandit Framework	The primary goal of this research is to reduce unnecessary handovers in an ultra-dense mmWave cellular network. Handover failure delays are not explicitly supported.
[42]	Ultra-Dense HetNet, mmWave, Multi-Connectivity/Dual-Connectivity	The authors proposed a dual connection-based handover scheme for an ultra-dense mmWave network to reduce frequent handovers by leveraging signal-to-interference-plus-noise ratios (SINR).

1.2 Problem Statement

Several mobility/handover management algorithms and schemes have been proposed for heterogeneous wireless networks. The next generation of mobile networks will be characterised by a massive deployment of heterogeneous base stations (BS). The high frequencies used in 5G cells become more challenging during user mobility since it will cause a high increase in handover failures, handover delay and inter-cell interferences. When tackling handover (HO) in 5G NR (New Radio), current mobility management solutions are generally divided at addressing the reduction of Handover Failures (HOFs), reduction of Unnecessary HOs and/or HO Ping-Pong (HOPP), or a combination of both, improving throughput or Quality of Experience (QoE) of the user and /or load balancing and reduction of HO signalling overhead. Other works are devoted to reducing HO Delay and Radio Link Failure (RLF). From the preliminary phase of this research, it was discovered that the

majority of the research works do not have mechanisms to minimise handover failures due to requisite resources not being available on the target base station. Hence, an efficient resource management scheme that minimises the dropping of handover requests due to resources not being available is of utmost importance to the problem of handover failure in mobility/handover management.

1.3 Research Objectives

This research looks at mobility management in 5G heterogeneous networks that use multiple Radio Access Technologies (RATs). The overarching purpose of this research is to create and test a mobility management scheme that ensures service continuity for a UE traversing a 5G network comprised of heterogeneous small cells. The following are the research's specific objectives:

- Examine the current state of the art in 5G mobility management and provide a critical examination of current proposals for alternative techniques in the context of 5G heterogeneous networks.
- Identify and analyse the existing gaps and limitations in current 5G handover schemes designed to mitigate handover failures.
- Design and model a handover scheme for heterogeneous 5G networks that aims to improve Handover Failure (HOF).

1.4 Research Questions

The research questions for this study from the overarching objective of this research which is to create and test a mobility management scheme that ensures service continuity for a UE traversing a 5G network comprised of heterogeneous small cells are:

- What are the key challenges and opportunities in 5G mobility management within heterogeneous networks, and how do emerging alternative techniques compare to existing methods in addressing these issues?
- What are the specific gaps and limitations in the current 5G handover schemes, and how do these factors contribute to handover failures in heterogeneous 5G networks?
- How can a novel handover scheme be designed and modelled to effectively reduce Handover Failure (HOF) in heterogeneous 5G networks while considering factors such as mobility patterns, network topology, and Quality of Service (QoS) requirements?

1.5 Scope and Limitations

1.5.1 Scope

This study deals with the issue of mobility management in 5G heterogeneous networks. The investigation specifically considers a scenario where users are connected to one base station (RAT-1), and it is moving toward another base station which is equipped with a different radio access technology (RAT-2). An admission control algorithm is also developed to minimise handover failure due to limited resources by utilising a semi-Markov handover prediction algorithm to prepare resources prior to handover execution. The

handover failure is evaluated by the number of handover calls dropped due to limited resources.

1.5.2 Limitations

The study uses an ideal free space propagation model with no obstacles. Radio waves propagate along with the line-of-sight path, which is described as the distance between the transmitting and receiving antennas and is also referred to as the communication range. The mobility of the user equipment is modelled using the random waypoint model, which is a simple, synthetic mobility model aimed at modelling intentional human movement in the range from 0.2 m/s to 2.2 m/s. The base station selection criteria is also based on the classical M/M/c queuing model.

1.6 Dissertation Outline

The rest of the dissertation is ordered as follows:

- **Chapter 2:** Review of the Literature and Background. This chapter provides a review of relevant literary works, as well as presents and discusses critical technical concepts relevant to this study, such as mobility management in mobile networks and handover management in 5G networks.
- **Chapter 3:** Heterogeneous Networks Mobility Management Scheme. The design of the system model used to investigate the research problem is presented and discussed in this chapter. In addition, the flow chart and the algorithm proposed to manage the handover from one RAT to another as well as allocating requisite resources are presented and tested in MATLAB.
- **Chapter 4:** Simulation and Results. The setup of the simulation environment, as well as the presentation and analysis of the results, are covered in this chapter.
- **Chapter 5:** Conclusion and Recommendations. This chapter draws conclusions and makes future work recommendations.

Chapter 2

Literature Review

2.1 Background

The International Telecommunication Union (ITU) has been directing government administrations, the telecommunication industry, and private sector efforts in the expansion of the global broadband international mobile telecommunication system (IMT). The IMT covers the global standards for wireless data networks starting from IMT-2000 (which was meant to support data transmission rates of 2 Mbps for fixed services and 384 Kbps in relation to mobile services; natively used in 3G standards), IMT-Advanced (is designed to provide a high-speed connection to fixed and mobile customers at rates ranging from 100 Mbps to 1 Gbps); 4G LTE and Worldwide Interoperability for Microwave Access (WiMAX) are native technologies, and recently IMT-2020 for 5G New Radio (NR). The 3rd Generation Partnership Project (3GPP) set necessities for the advancement of 5G systems, and it is accountable for creating and standardising, in addition to commercialising 5G verticals. The 3GPP standards body has approved a set of minimum specifications for 5G communications in both Non- Standalone (NSA) and Standalone (SA) operations. The first phase of 5G deployment, which is NSA focuses on enhanced mobile broadband (eMBB) which is geared for greater data bandwidth while running on both 4G LTE and 5G NR core networks. massive Machine-Type Communications (mMTC), as well as ultra-reliable, low-latency communications (URLLC), will also be native use cases for 5G [36]. eMBB will assist in accelerating applications like Augmented Reality (AR) and Virtual Reality (VR). mMTC supports a high density of devices with low data rates leveraging Machine to Machine (M2M) low cost and low power consumption benefits. URLLC supports safety-critical and mission-critical applications and technologies like self-driving cars, manufacturing automation and remote-controlled surgery which will require the full capabilities of 5G NR standalone deployment. The essential improvements of 5G NR (IMT-2020) capabilities are depicted in Figure 2.1 below. To satisfy these brand-new specifications, current network systems may be made denser by deploying small cells to form heterogeneous access networks (HetNets).

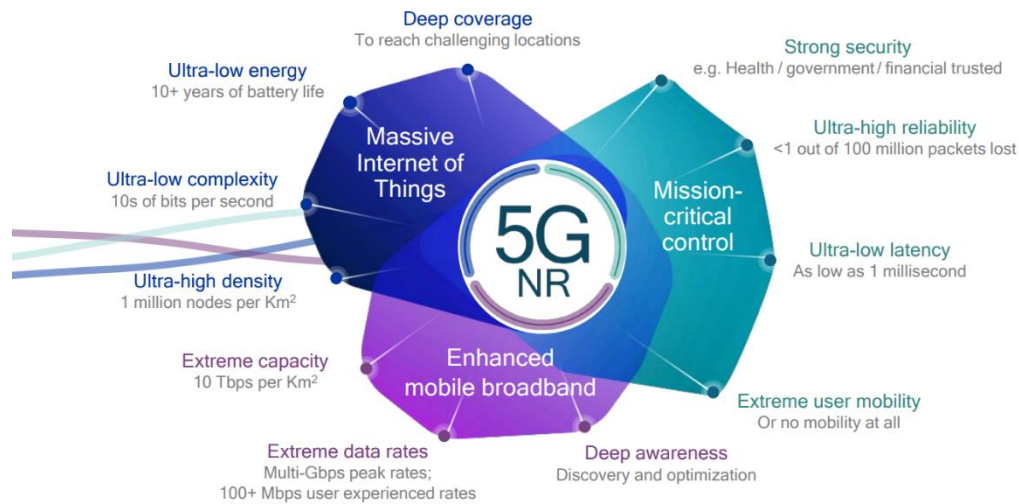


Figure 2.1 5G Usage Scenarios and Key Capabilities [57]

It is reported that at around 1947 at Bell Laboratories the history of cellular technology began with a technical report in which Rae Young proposed the now renowned hexagonal cell idea for a cellular mobile telephone network. It was later that year that Rae Young's ideas were expanded by his co-worker, Douglas H. Ring. Twenty years later, in 1969, Bell Labs implemented the first commercial cellular radio system materialising the ideas of Ring and Young. In this first cellular system, passengers of the *Metroliner* were able to make phone calls while travelling at 160 km/h. The train line was divided into nine radio zones with one base station each. As the train entered a new radio zone, any ongoing call was automatically transferred to the next base station [3].

The idea of transferring a call between cells for a moving mobile was further discussed in the articles published by Richard H. Frenkiel and Philip T. Porter in 1970 and 1971, respectively. Nonetheless, the procedures to carry out such a mechanism did not come to light until Amos E. Joel Jr published the very first patent on handovers in 1972 [4]. With this, the very first idea of providing service continuity for moving users was finally implemented in all subsequent cellular network generations and the concepts of *handover* and *mobility management* were born.

2.1.1 Mobility Management in Wireless Networks

First and second-generation network technologies are founded on the non-internet protocol (IP) based or circuit-switched setup. In contrast, next-generation (NG) (3G and beyond) wireless networks are based on IP (packet-switched) infrastructure. IP-built systems are appropriate for the swiftly growing mobile data and multimedia service industries because they employ the Internet model to telecommunications networks. Additionally, the internet paradigm offers an improved solution to integrate heterogeneous radio access technologies (RAT) seamlessly [5].

Presently, different wireless access standards that cater for various access requirements and profiles of mobile users are in existence as depicted in Figure 2.2 below: for wireless communications over a short distance, Wireless Local Area Networks (WLANs) are acceptable deployment solutions. For coverage over a large geographical region, conventional, current, and future cellular networks, known as, next-generation (NG) and fibre offer multimedia solutions. For global coverage, satellite systems have been deployed expansively for private and commercial applications. These networks and a host of others

are complementary in nature; hence their integration allows users to join the best accessible network that fulfils its requirements [38].

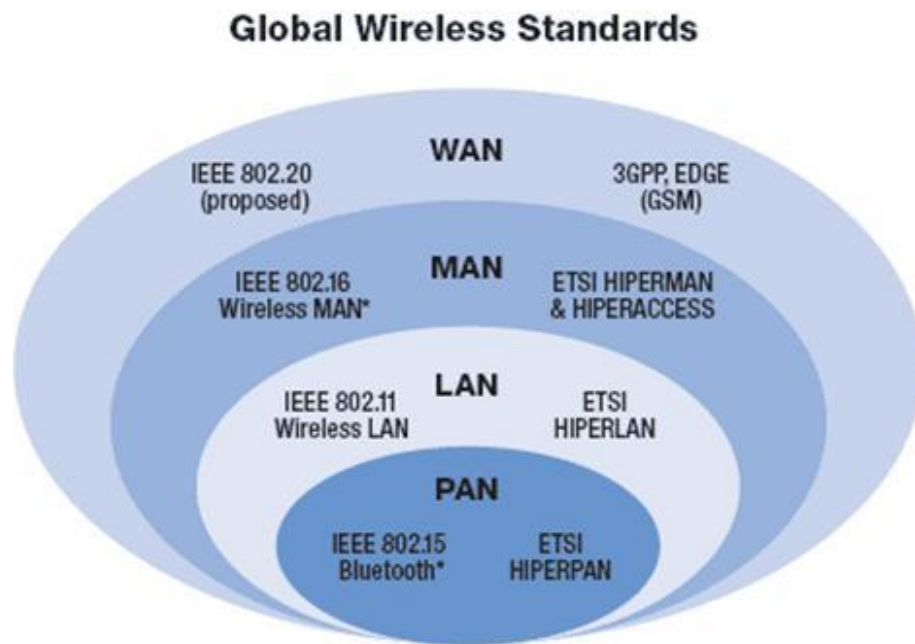


Figure 2.2 Classifications of Global Wireless Technologies [43]

Table 2.1 Classifications of Global Wireless Technologies

	PAN	LAN	MAN	WAN
Description	A Personal Area Network	Local Area Network	Metropolitan Area Network	Metropolitan Area Network
Connection Standards	Bluetooth	802.11	802.11 WiMAX	GSM, GPRS, HSDPA
Speed	<1Mbps	11 to 54 Mbps	11 to 100+ Mbps	10 to 384 Kbps 1.8/3.6 – 7.2 Mbps
Propagation Range	Short	Medium	Medium and Long	Long
Applications	Peer-to-Peer Device-to-Device	Enterprise Networks	Last mile access	Mobile Phone, Cellular data

2.1.2 Mobility Management in 4G

The architecture of the LTE network comprises evolved NodeBs (eNodeBs/eNBs) offering a user plane (UP) and control plane (CP). The eNodeBs are connected by way of an X2 interface. The core function of the Serving Gateway (S-GW) is steering and forwarding mobile user data packets. The S-GW is also accountable for handovers between different eNBs (inter-eNB handovers) and offers mobility between LTE and other cellular networks, like 3G. The Mobility Management Entity (MME) is a critical LTE control node for managing subscriber authentication, establishing bearer path, bearer activation/deactivation, roaming and handovers to other networks [10].

The 4G mobility management comprises extra mobility related functionalities, which were not available in previous networks. The new examples include moving networks, seamless roaming, and vertical handover. The UE mobility management states in 4G LTE are classified into Idle mode and Connected mode. In the idle mode, mobility depends on UE measurements and UE decisions on whether to make a cell selection or reselection and the MME keeps track of UEs that are registered in the Tracking Area (TA). When a phone call is made, the idle mode is switched to the connected mode. In connected mode, mobility relies on UE measurement report and eNB decisions to make a handover, and the eNB keeps track of the UE [11] [12]. The LTE handover procedure, including the three major handover phases, is shown in Figure 2.3. The handover procedure consists of about 5 parts: **Initiation** phase, **Preparation** phase, **Decision** phase, **Execution** phase and **Completion** phase.

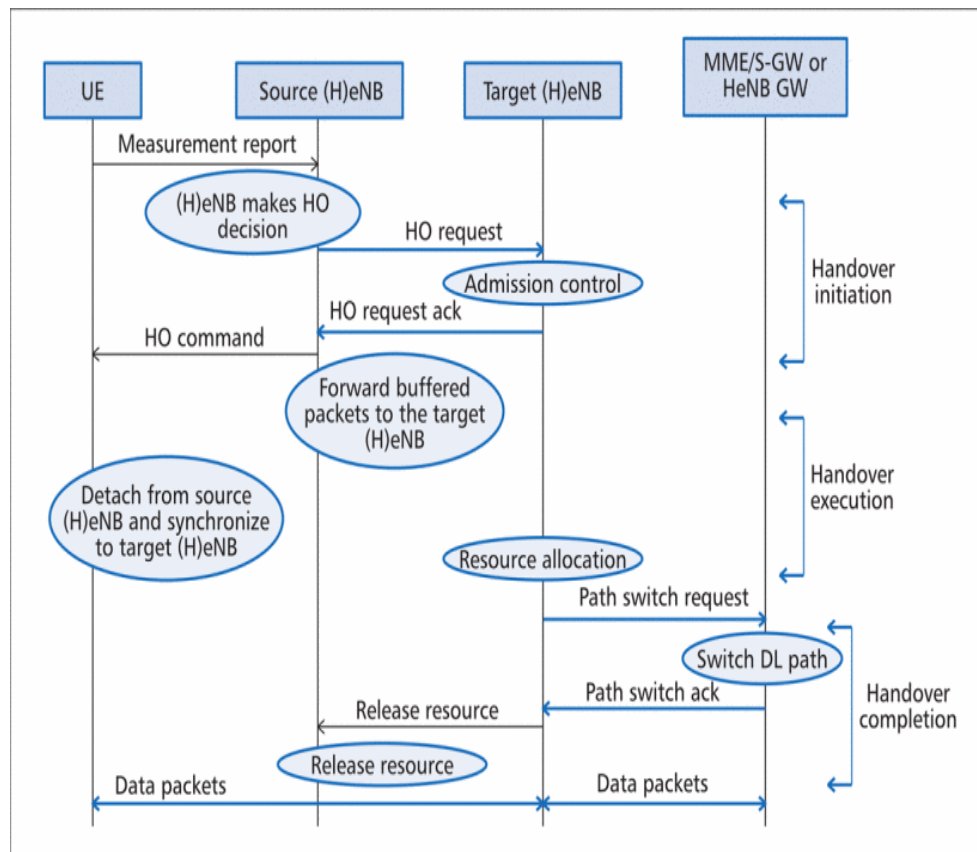


Figure 2.3 LTE Handover [61]

2.1.3 Towards 5G Mobile Network

New features and mobility enhancements have appeared since the first version of LTE by the 3rd Generation Partnership Project (3GPP) Release 8. LTE-Advanced was standardised in Release 10, and it included features like *Carrier Aggregation* (CA). CA is a technique to increase the data rate per user, where numerous carrier frequency blocks are allocated to the same user within the same base station site. This functionality was later extended to the aggregation of different carriers belonging to different sites in Release 12 and dubbed *Dual Connectivity* (DC). Dual Connectivity, like Carrier Aggregation, aims to increase the throughput of the UE by utilising resources from multiple carriers as well as to attain improved mobility robustness [13]. In DC, UE can concurrently be associated with two cellular groups referred to as a master cell group (MCG) and secondary cell group (SCG) over master eNB (MeNB) and secondary eNB (SeNB), correspondingly. The DC technique offers two concurrent data paths (user plane connections) on diverse frequencies like a Macro cell (4G) and a small cell (5G) for a UE. Even though it increases data throughput, dual connectivity comes with the price of numerous mobility occurrences [22]. Apart from normal handovers, new mobility events are specified for the aggregation, substitution, and release of the cells that serve the additional radio links [32].

The LTE standard continues evolving from the new features and enhancements included in LTE-A to the most recent Release 13 under the commercial name of LTE-A Pro. However, the continuous demand for higher data rates and the appearance of new use cases (ranging from ultra-high-definition media content to safety and ultra-low-latency applications), calls for new capabilities that current mobile networks are unable to provide. The 5G new radio (NR) architecture, promises an entirely new design paradigm that will cater for more rigorous and demanding conditions, paving the way for the employment of heterogeneous and integrated networks and technologies. 5G will rely heavily on software and virtualisation. To achieve the desired results, flexibility, agility, and scalability, 5G will be driven by technologies in development, such as Network Slicing, Software Defined Networking (SDN), and Network Functions Virtualization (NFV), Fog Computing (FC) and Multi-Access Edge Computing (MEC). Other key features envisioned for 5G/Beyond 5G (B5G) are Spectrum aggregation, Dense/Ultra-Dense cells (Small cells), Spatial randomness of network tessellation, Multiple Input Multiple Output (MIMO) antenna systems, mmWave, Seamless handovers for high-speed trains moving at ultra-high speeds of up to 500 km/h. Table 2.2 shows a comparison between the current design recommendations for LTE-Advanced and the new target specifications for 5G.

Table 2.2 Comparison between the Mobility Requirements for 4G and 5G [39]

Parameter	4G	5G
Handover Interruption Time	27.5 ms [Intra Frequency HO] 40-60 ms [Inter Frequency HO]	0 ms (Target for 3GPP)
Operating Frequency	≤ 6 GHz	≤ 6 GHz, 28 GHz, 39 GHz, up to 52 GHz
Carrier Aggregation	≤ 32	≤ 16
Bandwidth support	≤ 20 MHz	≤ 100 MHz (6 GHz), ≤ 1 GHz (> 6 GHz)
Analog Beamforming	N/A	Supported
Digital Beamforming	Up to 8 layers	Up to 12 layers
Subcarrier Spacing	15 kHz (fixed)	Ranges [15 kHz - 240 kHz]
Maximum UE Speed	350 k/h	500 k/h

2.1.4 Handover Management in 5G

Handover is a procedure of changing the user equipment (UE) point of attachment in respect of frequency, time slot, spreading code, or a combination thereof while the UE is in motion so that the best serving base station (S-BS) is constantly chosen while the user equipment is still in a connected mode. The traditional method for selecting the best S-BS usually depends on the intensity of the received signal (RSS) of base stations near the UE. When a user is in motion, the UE may want to change its point of attachment if another BS offers a higher RSS than the serving BS. The handover management literature lists numerous factors that influence handover management, such as user choices, network selection strategy, mobility situations, mobility protocols, handover controls, and so on.

A Next Generation Radio Access Network (NG-RAN) node in a 5G network is any of a gNB that provides NR access to the UE or an ng-eNB that provides E-UTRAN access to the UE [19]. In 5G RAN, a UE device in RRC-CONNECTED mode can use multi-RAT dual connectivity, i.e., multi-connectivity operation. Multi-connectivity in 5G can be construed as an advancement of LTE Dual Connectivity [20]. Multi-connectivity [23] is a 3GPP-specified protocol that reduces the number of handover failures in a heterogeneous network (HetNet). This protocol operates in such a way that a UE is simultaneously linked to two 3GPP Radio Access Network (RAN) nodes with various zones of coverage, such as a macro cell and a femto cell. Data for a specific protocol data unit (PDU) session is forwarded to UE via two distinct gNBs. Multi-connectivity enhances 5G network performance for various use cases. The benefits of the multi-connectivity protocol include increased mobility robustness, reduced handover interruption time (HIT), and increased network overall reliability. Schemes and mechanisms that use multi-connectivity protocol are ITU-Vertical Multihoming (ITU-VMH), Co-ordinated Multipoint (CoMP) strategy and Dual Connectivity (DC). 5G presented an additional third state for a UE, as compared to the two states of LTE, termed INACTIVE. The significance of the INACTIVE state is the reduction of time taken to switch to the CONNECTED state [19], figure 2.4 below.

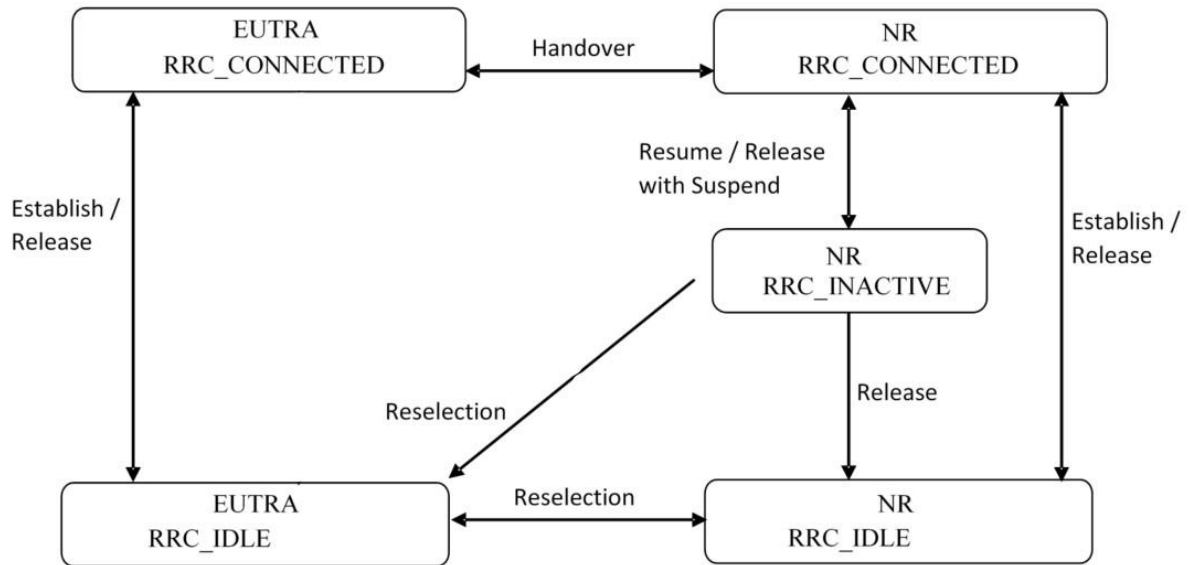


Figure 2.4 UE state machine and state transitions between NR/5GC, E-UTRA/EPC and E-UTRA/5GC

When an RRC (Radio Resource Control) link connection is established, a UE is either in the RRC CONNECTED or RRC INACTIVE state. If the opposite occurs, i.e., no RRC connection is formed, the UE enters the RRC IDLE state. When in RRC connected mode, to start measurement and make a handover, the procedure is done by actions called “Event”. The several types of events are listed and explained in Table 2.3 below.

Table 2.3 Handover Measurement Report Triggering Events

Event	Entering Situation
A1	The signal of the serving cell exceeds the threshold in strength.
A2	The signal of the serving cell falls below the threshold.
A3	The signal of the neighbour cell becomes stronger than the signal of the Serving cell by a margin of the offset value.
A4	The signal from a neighbouring cell becomes stronger than the threshold.
A5	The signal from the serving cell becomes weaker than threshold 1, while the signal from the neighbouring cell becomes stronger than threshold 2.

2.1.4.1 5G Baseline Handover

The baseline HO procedure in 5G is shown in Figure 2.5 below, containing Handover Preparation (0-5), Handover Execution (6-8) and Handover Completion (9-12) stages [21]:

- ❖ **Step 1:** The serving gNB receives a Measurement Report (MR) from UE. UE measures the signal strength of the serving gNB (S-gNB) and neighbouring gNBs (N-gNBs) and determines whether the measured signal strength and other parameters satisfy the HO event's entering condition. See Table 2.3.
- ❖ **Step 2:** The serving gNB makes a HO decision on the basis of MR and RRM information.

- ❖ **Step 3:** The serving gNB sends the target gNB a HO REQUEST message.
- ❖ **Step 4:** If resources are available, Target gNB will run the Admission Control procedure.
- ❖ **Step 5:** HO Request Acknowledgement by the destination gNB and initiation of data forwarding.
- ❖ **Step 6:** A HO command is issued to the UE. from source gNB.
- ❖ **Step 7:** The serving gNB sends a transfer message to the destination gNB.
- ❖ **Step 8:** The session with the serving cell is terminated, and communication with the new cell is initiated.
- ❖ **Step 9:** AMF Path Switch Request message sent with notification of changes.
- ❖ **Step 10:** Downlink data path switch
- ❖ **Step 11:** The AMF accepts the Path Switch Request.
- ❖ **Step 12:** HO success and resources released.

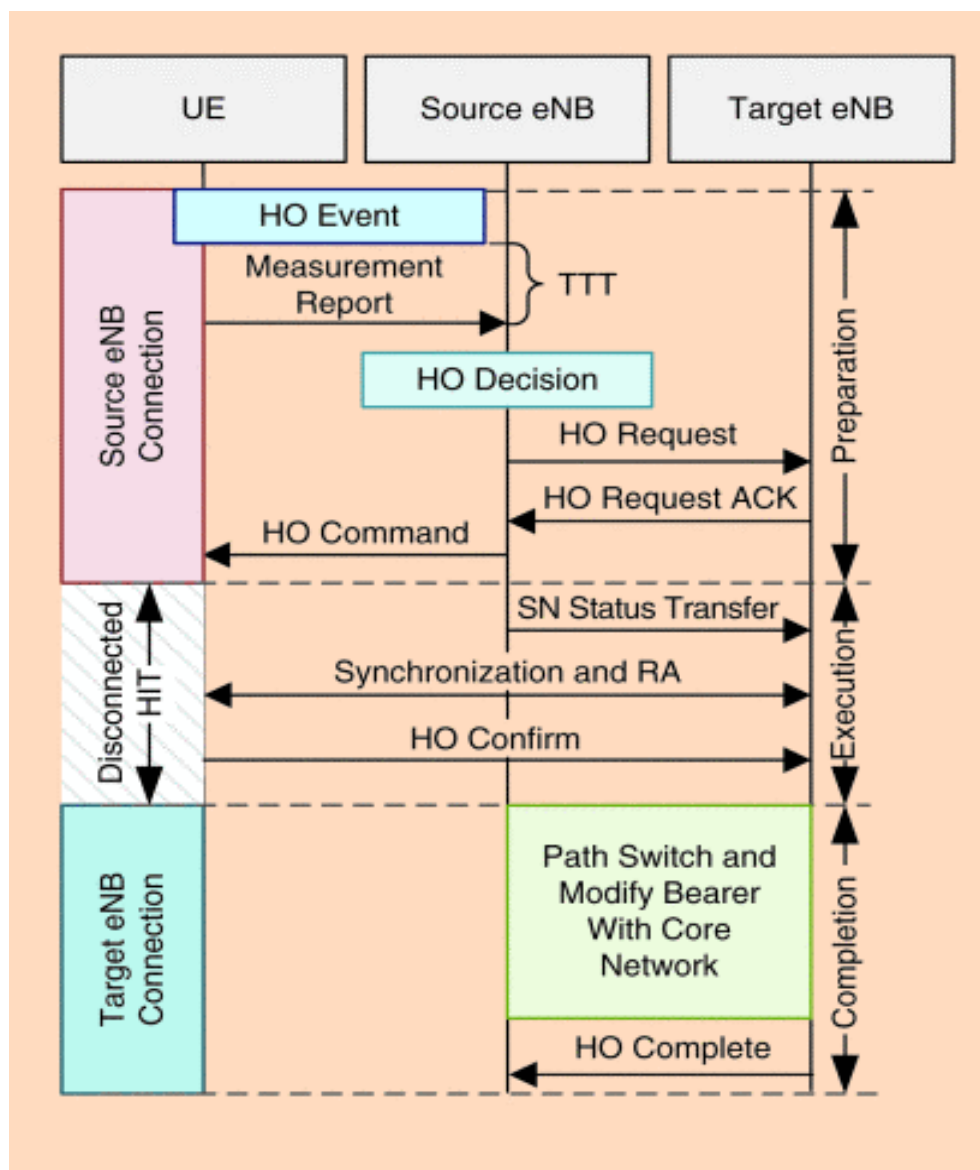


Figure 2.5 The Baseline HO Procedure in 5G [31]

There are two varieties of handovers.: intrasystem/horizontal handover and intersystem/vertical handover. A horizontal handover within homogeneous networks occurs

as the strength of the serving BS's signal falls beneath a prior set threshold. A vertical handover amongst heterogeneous networks may result due to the subsequent circumstances (Mohanty, 2006) – (i) when a UE moves away from the serving cell and enters the territory of an overlying network, (ii) when a user attached to a network prefers to handover to an underlying or overlaid network for their service requirements, and (iii) when load balancing among adjacent networks is required. With the increasing heterogeneity of cellular networks, the following metrics may also be used to select the best serving base station: load balancing, delay, and throughput.

Handover can also be allocated into two other broad groups: hard handover and soft handover. They can also further be grouped by “break before make” (BBM) and “make before break” (MMB). With respect to hard handovers aka BBM, existing radio resources are freed before brand new resources are acquired, whereas, in soft handovers (MMB), both existing and new radio resources are utilised during the handover process. [18].

From Figure 2.6 [19], subsequent HO types in 5G NR are defined:

- ❖ **Intra-gNB HO:** Serving and target cells are served by the same gNB, Figure 2.6 (a).
- ❖ **Inter-gNB HO:** Serving cell and target cell are served by different gNBs. The assumption is that the AMF will not be changed due to the occurrence of a HO. Furthermore, the UPF may or may not be repositioned, Figure 2.6 (c) and Figure 2.6 (b) correspondingly.
- ❖ **Inter-gNB HO with AMF Change:** At this stage, the handover includes a variation of AMF. In addition, the UPF may or may not be displaced as depicted in Figure 2.6 (e) and Figure 2.6 (d) correspondingly.

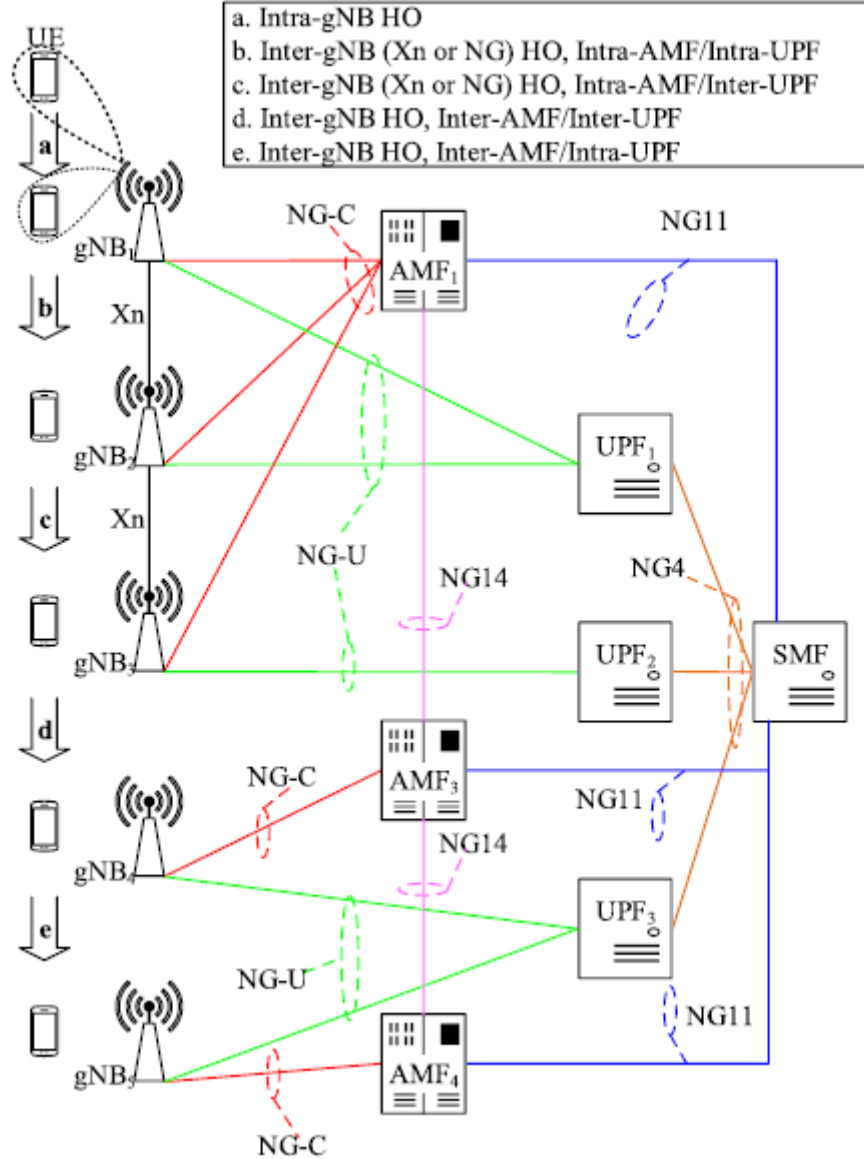


Figure 2.6 5G NR Mobility Architecture and HO Use Cases [19]

2.1.4.2 5G Conditional Handover (CHO)

Conditional handover (CHO) was proposed by 3GPP in Release 16 as a mobility enhancement to expand 5G baseline HO's reliability, focusing on reducing the number of HO Failure (HOF) occurrences. CHO can accomplish a superior HO achievement degree by executing an early handover preparation when the current connection is still steady. Figure 2.7 [31] shows the 5G CHO procedure. According to Lee, Changsung, et al., there are two autonomous HO triggering events, described as the preparation event (*PEvent*) and the execution event (*EEvent*) with respect to CHO. According to [31], current CHO studies consider mostly events (Table 2.3 above) A3 and A4 for the *PEvent*, while A3 and A5 are used for the *EEvent*. In conditional handover, several potential target cells are prepared beforehand, which allows the handover command to be sent to the user terminal in advance while radio conditions have not deteriorated.

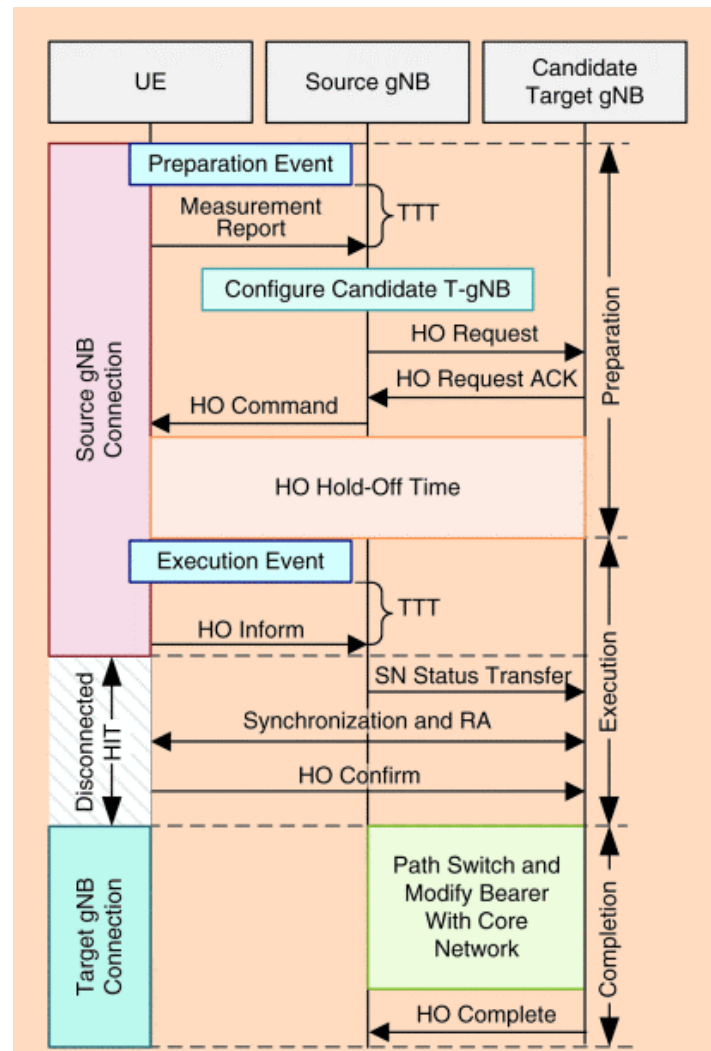


Figure 2.7 The 3GPP 5G Conditional Handover Procedure [31]

2.2 Related Works

In this section, a critical review of selected mobility management schemes is presented. The schemes and algorithms are classified into Software Defined Networking (SDN) and/ or Network Functions Virtualisation (NFV)-Based Mobility Management Solutions, Mobility Management Solutions in Sliced Environments and State-of-the-Art Mobility Management Solutions. Traditional handover solutions base the handover decision solely on the Received Signal Strength (RSS) level and an additional metric to support the handover procedure, despite being uninvolved in the handover decision process. The conventional handover algorithms are more suitable for horizontal handover networks.

2.2.1 SDN and /NFV – Based Solutions

With the spread of network gadgets like smartphones and tablets, novel data applications and use cases, wireless systems are seeing an astronomical surge in traffic volume. As a result, the novel fifth-generation wireless network ecosystem is anticipated to be densified with heterogeneous small cells. Consequently, the deployment of heterogeneous cells comes with the challenge of frequent and unnecessary handovers in

addition to the Handover Ping Pong (HOPP) effect. Ping-pong handover refers to frequent handover to and from a cell pair. The ping-pong effect is caused by mobile units moving back and forth between cell pairs or by high signal fluctuations at the cell pair's common boundary. But the emergence of technologies like SDN and NFV promises to offer solutions to the above challenges as well as a host of others by providing network programmability, flexibility, and simple management.

SDN and NFV are two distinct technologies that complement one another. The SDN controller (control plane) offers an overview of the whole network of the data plane, so some functionalities like handover control can be executed in the SDN architecture to efficiently complete the handover process. In addition, SDN grants mobile network operators reduced CAPEX and OPEX and applications and services can be instantiated in less time. NFV simplifies network functionality by decoupling network functions from proprietary hardware appliances by running them on standard servers in the data centre. SDN and/or NFV mobility management schemes or algorithms can take different, or a combination of forms depicted in Figure 2.8 below.

The authors in [25] proposed a semi-centralized Vertical handover technique based on SDN for handovers from an LTE eNodeB (Macro Cell) to a 5G Femto cell. The scheme is based on a combination of criteria, including network context data (for example RSS, BER, available bandwidth, delay, and jitter), data on the user's context (for example, user speed), and details about the service context (for example type of service). The emphasis is on SDN's role in the handover (HO) decision phase, where the Quality of Service (QoS) function is used to compute a network score. In the SDN controller, the score of three separate networks is dynamically computed, and the network with the highest score is picked. To avoid unnecessary handovers, UE speed is considered as well. SDN is used to help improve the HO procedure in relation to handover frequency, signalling overhead or cost, HO delay, and QoS. In essence, HO signalling messages are decreased since the HO messages exchanged are minimised. The SDN Controller has a global view of the sub-network and is in charge of collecting data from the network prior to making a HO decision, thereby lowering the HO delay. According to [33], the device-to-device (D2D) strategy for handover demonstrated that it could improve handover performance in terms of the number of signalling messages, handover latency, and handover rate when compared to current schemes. The D2D handover decision method and management in the SDN-based LTE-A/5G architecture maintain side link stability. A D2D communication handover procedure was presented for assessing handover effectiveness metrics, for example, data transfer latency. Recently, the Internet Engineering Task Force (IETF) charted a network-based technique known as the distributed mobility management (DMM) protocol, which aims to flatten the network architecture in order to address the mobile Internet's rampant scalability and reliability shortcomings. A method based on SDN to provide DMM is among the candidate solutions [34][35] because DMM still carries a hefty price of overhead, excessive delay during handover, high packet loss, and tunnel management crisis. The approach in [34] develops the location and handover management features within a centralised SDN Controller whereas the data steering functionality is entirely dispersed within the access routers. However, the authors only consider static mobility management without considering the QoS requirements on the flows. In [35], the authors proposed an SDN-based fast handover with an optimum routing control scheme for DMM to boost the handover performance of the mobility anchor functionality over the DMM. An analytical evaluation and performance analysis are carried out for signalling cost, handover latency, throughput, and packet loss. However, it does not consider load balancing and has a high handover

request dropping because of increasing attempted handovers.

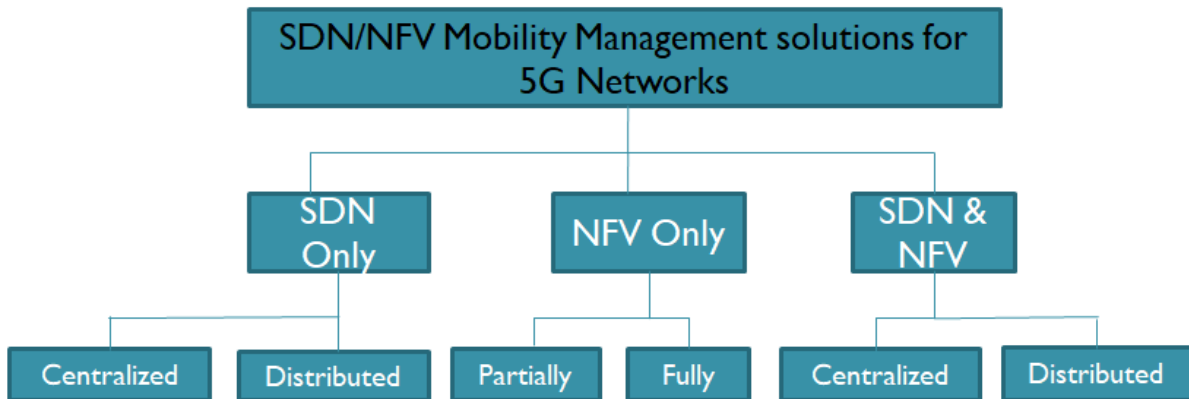


Figure 2.8 Classification of SDN and NFV Mobility Solutions

2.2.2 Mobility Management Solutions in Sliced Environments

Mobility management in a 5G sliced environment presents unprecedented and complicated challenges as devices/user mobility should also be managed for different slices. To that end, new mobility solutions need to be standardised for such scenarios. The ability to support slicing is among the essential prerequisites of the 5G networks. Multiple logical networks can be overlaid on top of a single physical infrastructure shared by all using network slicing. Each slice of the network is completely separated and independent, which enables new slices to be added without impacting the rest of the network. Network Slicing enables simultaneous provisioning of different use cases, for example, massive Machine Type Communications (mMTC), Ultra-Reliable Low Latency Communications (URLLC) and enhanced Mobile Broadband (eMBB).

Inter-slice mobility management solutions need to observe the standard building framework of the network slicing agenda specified by 3GPP, according to a Service-based Architecture (SBA) model. The network slicing framework, though, does not provide any specific sustenance for inter-slice handover; hence, whenever a UE needs to switch from its current slice, its current data session is terminated before it can be re-established to another qualifying slice.

In [24], a modification to the 3GPP multi-Connectivity protocol (Release-15 5G) and LTE-WAN Aggregation (LWA Release 13-LTE) to enable UE slice mobility to two or more base stations is proposed. The base stations can include femto-femto, macro-macro and femto-macro. The improved procedure also allows for multi-connection to nodes from several RATs, such as NR gNB and Wi-Fi AP, as well as data transfer over multiple radio connections to a single UE. Furthermore, the proposed enhancement also considers usage information per base station like load, interference, etc. for user association. Another benefit of the 3GP multi-connectivity slice deployment is the increased flexibility, as all slices no longer need to be installed on all radio nodes within the Registration Area (RA). The authors also claimed that the proposed enhancement can be utilised for slice handover through a large contingent of base stations for immobile users.

The authors of [29] propose a line of work in which the main concern is to investigate the inter-slice handover in the V2X (Vehicle to Everything) tele-operated driving leveraging SDN, by developing an admission control mechanism with an effective resource management strategy to look for the accessibility of radio resources in the radio access network (RAN) to accept handover demands. If no resources are available in a slice to admit

the request, effective resource management should be implemented between slices prior to the handover. The authors concentrate on inter-slice handover management when the target slice resource is full. To accept the handover request, a resource borrowing mechanism is used to source resources from other accessible slices. The primary goal of this work is to lessen the likelihood of failure of handover in the overloaded slice while adhering to the limitations detailed below: The handover session's quality of service, careful consideration is observed about the load constraint of the lender slice such that it is not violated, the amount of offered resources borrowed or shared between slices should be fair and handover failure minimised at the lender slice. The constraints in themselves have contradictory concerns. Hence, to accomplish the ultimate performance, a bargain is considered. To that end, the authors devised two distinct formulations of bargaining games. Even though their method increases the delay of handover, it still achieves an acceptable level (12 ms) for a high arrival rate and does not violate URLLC obligations.

2.2.3 State-Of-The-Art (SOTA) in 5G Mobility Management Solutions

Jain et al. [37] noted that RAN technologies are anticipated to undertake a crucial revolution in the next-generation networks. Such technologies may include Visible Light Communication (VLC), Light Fidelity (LiFi), etc. Due to the high frequency of VLC and LiFi, and TeraHertz (THz) band technologies, they are significantly susceptible to environmental impacts. Consequently, that affects QoS and introduces new complex mobility patterns due to additional use cases involved.

Consequently, a new perspective, in which Mobility Management solutions are distributed and elastic, can accommodate manifold use cases concurrently and cater for other sweeping variations in 5G and beyond 5G (B5G) systems with reliability, is vital. Flexibility will allow novel mobility management solutions to adjust to the user and network context. Moreover, reliability is going to assist in providing seamless mobility.

The authors of [26] extended their previous auto-tuning optimisation algorithm, which uses user velocity and Received Signal Reference Power (RSRP) to adjust the HO Margin (HOM) and time to trigger (TTT). They suggest a velocity-based self-optimisation algorithm to regulate the HO control parameters in 4G/5G HetNets and address the issue of Mobility Robustness Optimization (MRO) by concentrating on two performance metrics: radio link failure (RLF) and handover ping pong (HOPP). Upon evaluation, the scheme attained a low call drop rate and reduced HO delay as well as a reduction in HO interruption time. The proposed algorithm does not consider load balancing and over-simplified and unrealistic simulation scenarios are assumed like UE moving with the same speed and utilising the no-restrictions line of sight (LOS) connections, and it is based on a meta-heuristic algorithm, in which the solution is not globally optimal but a sub-optimal solution with low computation complexity.

Paper [27] presents a scheme that anticipates congestion and incorporates a load-aware association technique for maintaining the conflict-free operation of load balancing and capacity and coverage optimisation self-organizing network (CCO SON) functionalities code-named "OPERA." The authors propose a framework for dynamic network orchestration based on mobility prediction in order to achieve load balancing with QoS constraints. It makes use of a semi-Markov model to forecast spatial-temporal mobility in conjunction with the vector theory geomaker notion to anticipate future cell loads in HetNets. Then, based on the predicted cell utilization, proactive optimisation is carried out in order to maximise the logarithmic sum of all cells' free resources.

References [28], [30], and [31] propose handover schemes in 5G mmWave band

cellular networks by leveraging Artificial Intelligence (AI) and/or Machine Learning (ML) to predict the next base station based on empirical data from the user's post-handover trajectory. In [28], the authors enhanced the baseline X2 handover mechanism by implementing and adding an algorithm to modify it so that it can predict the next base station (BS) to improve UE QoS. A grid-like structure is considered for the network layout for urban cities BSs and UEs are arranged in a horizontal direction and in opposite, while some UEs are kept motionless and straight trajectories in the cartesian coordinate system assumed. Radio frequency indicators collected from both BS and UE include signal to noise ratio (SNR), Beam ID, and Timestamp. Simulation results showed that handover delays in horizontal, vertical, and vertical and horizontal combinations are reduced to a reasonable amount. Whereas the authors of [30] propose a novel mechanism to prevent needless handovers by selecting the next BS to which a user must pass over control so that the new connection may be established to continue for as long as possible. The decision on the next BS requires some knowledge of the user's post-handover mobility trajectory and Line of Sight (LOS) blockage, which is learned online using a multi-armed bandit framework. A numerical performance analysis is performed, which demonstrates the proposed mechanism for reducing handover in scenarios where user mobility follows predictable patterns. The downside of the mechanism is that it adds a lot of signalling and computational complexity because the multi-armed bandit (MAB) process keeps a cumulative reward computed by adding the awards obtained by all previous UEs who switched to the serving base station during previous handovers in that grid.

The paper in [31] presents a novel conditional handover (CHO) plan using deep-learning technology to forecast the future BS and overcome the shortcomings of baseline CHO in order to reduce Handover Failure (HOF). In the proposed scheme, event A2 is utilised to control the timing of the forecasting based on the channel standing of the serving gNB (S-gNB). Furthermore, event A3 is utilised as the handover execution event (EEvent) because the channel status of the target gNB (T-gNB) is required to ascertain the implementation of the handover. The scheme shows a promising reduction in signalling overheads caused by sending measurement reports (MR) and handover command (HCMD) messages.

Taking into consideration a two-tier dense HetNet deployment scenario in which macro and small cell BSs are distributed based on two autonomous homogeneous networks Poisson Point Processes (PPPs), the authors in [45] proposed an analytical model. Their model includes coverage models, representing the small cell boundary and the cross-tier handover failure boundary with two biased circles and stochastic geometry. Their simulation results showed that there is a gap compared to analytical results. This is because the authors assumed that during a single trace movement in the analysis, a standard user in motion crosses the small cell coverage. However, in the simulation, the user requires multiple trace movements to cross the coverage at times, which is not accounted for in the handover and handover failure counts. Table 2.4 below shows a qualitative analysis of the reviewed literary works.

Table 2.4 Overview and Qualitative Evaluation of Handover Schemes

Scheme	Scope	Environment	Signalling Cost	Load Balancing (QoS / QoE)	HOF	HO Delay
[24] 2019	3GPP Multi-connectivity Protocol and Slice Mobility on the RAN	HetNet		✓	✓	
[25] 2018	Semi-Centralised SDN- Based Vertical Handover Decision (VHD) scheme within 5G Femto- LTE eNodeB (Macro) HetNet considering network, user, and Service context information & QoS	5G HetNet	✓	✓		✓
[26] 2020	Velocity-Based Optimisation Algorithm that dynamically adjusts HCPs (TTT & HOM) in accordance with UE Speed and RSRP	4G/5G HetNets	✓		✓	✓
[27] 2020	A proactive dynamic network orchestration framework for load balancing with QoS constraints for HetNets based on mobility prediction.	5G HetNet	✓	✓		✓
[28] 2019	A modification of the LTE baseline X2 Handover Mechanism with an algorithm that predicts the next base station using SNR indication for a UE moving in a straight trajectory either horizontally or vertically and a combination of both on a cartesian plane in 5G mmWave	5G mmWave		✓		✓
[29] 2020	Inter-slice handover management for the Automotive vertical using SDN by deriving an admission control with an efficient resource management scheme in a sliced environment when the target slice is overloaded by borrowing resources from other slices using bargaining games.	5G	✓	✓	✓	✓
[30] 2019	Handover mechanism in a 5G mmWave cellular network to reduce unnecessary handovers by carefully deciding the next base station (BS), so that the new user-BS connection after the handover can last for a comparatively longer time. The scheme leverages the empirical distribution of users' post-handover trajectory and LOS blockage, learned online via a multi-armed bandit (MAB) framework.	5G mmWave		✓		

[31] 2020	A novel Prediction-Based Conditional Handover (PCHO) scheme for 5G mmWave networks that use deep-learning technologies to predict the next gNB to overcome the shortcomings of baseline CHO.	5G mmWave	✓	✓	✓	
[33] 2020	To maintain side link stability, an improved D2D handover management in an SDN-based LTE-A/5G architecture is used.	LTE-A/5G HetNet	✓			✓
[34] 2017	A distributed mobility management (DMM) approach based on software-defined networking (SDN) for 5G networks.	5G	✓			✓
[35] 2019	For distributed mobility management (DMM), an SDN-based fast handover with the best routing control method is used.	5G	✓	✓		✓
[42] 2019	A dual connection-based handover scheme for ultra-dense mmWave networks is proposed.	5G HetNet		✓	✓	✓

Table 2.4 above evaluated recent works (2017 to 2020) in the space of 4G and/or 5G HetNets based on the reduction of four selected metrics, namely signalling cost, load balancing, handover failure and handover delay. A clear observation from the table is that majority of the works, even though most of them tried to accommodate at least two metrics, did not try to solve the phenomenon of handover failure in a heterogeneous environment. Those that incorporate reduction of handover failure do so in manner that utilises tried and tested 5G techniques such as dual connectivity, multi-connectivity protocol and conditional handover. It must be noted that the only exception to the above revelation is the work in [\[31\]](#) which incorporated the idea of prediction to enhance the baseline 5G conditional handover algorithm.

2.3 Chapter summary

This chapter is dedicated to reviewing literary works in the 5G mobility/handover management. The chapter started off with a brief background of mobile wireless networks from 3G to 5G, setting a tone for early deployment of 5G which utilises 4G core network together with 5G core. Thereafter a history of handover management set the stage for discussions on 4G handover management toward 5G handover management. The baseline handover architecture and procedure for 5G is discussed, which leads to a more relevant discussion of 5G conditional handover which seeks to reduce handover failure. Finally relevant literary works are reviewed, and a quantitative evaluation/comparison carried out to realise some gaps that still exist.

Chapter 3

5G Heterogeneous Networks Mobility Management Scheme

3.1 Heterogeneous Networks and Small cells

A system of heterogeneous 5G access networks (HetNets) integrates two diverse classes of cells with different access technologies, for example, macrocells (4G) and small cells (5G). A typical HetNet operation incorporates a combination of legacy network systems and modern radio access technologies (RATs) where legacy technologies are used as umbrella cells over the small modern cells. Figure 3.1 below shows the system architecture of a HetNet. The different types of cells in a HetNet are elaborated below.

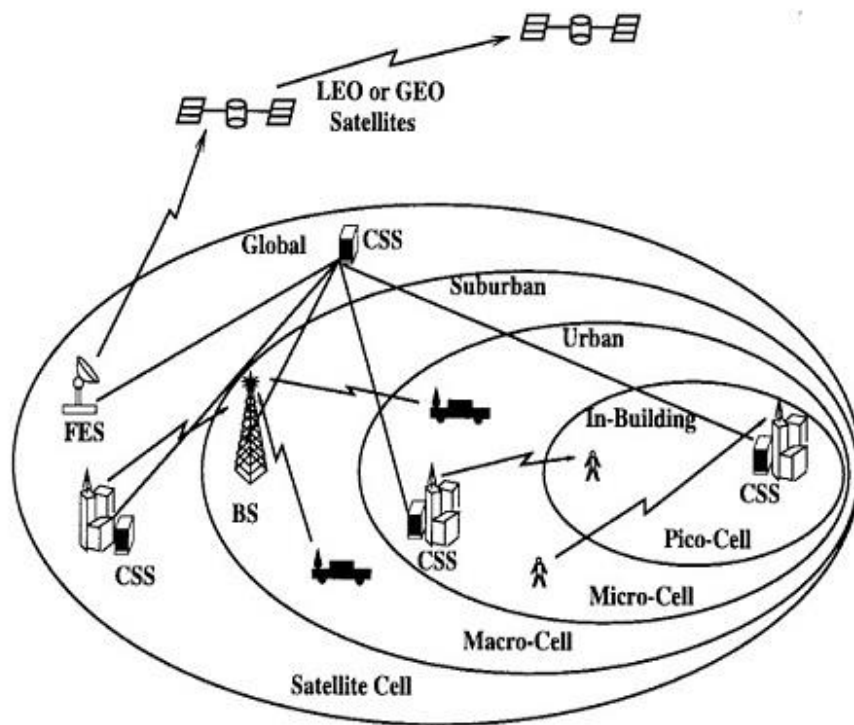


Figure 3.1 System Architecture of Intra-system and Inter-system Roaming in Different Types of HetNets

3.1.1.1 Femto Cell

Femtocells are low-power, small cell stations designed to work in the licensed indoor spectrum for use in residential or corporate settings. The transmit power of femto cells is in the range of fewer than 100 megawatts and covers a distance in the range of fewer than 30 meters [25]. The femto cell redirect the communications traffic to the mobile core network over a fixed broadband network using a mobile handset [16].

3.1.1.2 Pico Cell

Picocells are small cellular base stations naturally used to augment capacity in populated areas, such as train stations, shopping malls, stadiums, and city centres. Their reduced size and lower operating costs make placement more versatile. The typical range of a picocell is smaller than a microcell (in both range and capacity), usually about 100 metres. They are suitable for indoor and outdoor communications as compared to Femto cells.

3.1.1.3 Micro Cell

Microcells are mini cell towers that can be affixed to already erected physical infrastructure (such as streetlights, bus shelters or rooftops of tall buildings) in areas where traditional cell towers simply cannot be utilised. The minimum coverage range of around 500 m [17] and the coverage range can span up to 2 kilometres. Although, if the microcell is utilising extremely high frequency ranges like mmWave, building structures and similar high rising objects will block the line of sight of the signal. To make up for capacity and coverage gaps, these cells may be set up either indoors or outdoors.

3.1.1.4 Macro Cell

Macrocells are typically used to provide blanket radio coverage of a large geographical area of mobile network access served by high power cell sites (tower, antenna, or mast). Macro cells are used for blanket coverage whilst minimising initial cost of deployment. Typically, macro-cellular networks have cell sizes of 1 kilometre to 20 km.

Table 3.1 Comparison of Small Cells

	Output Power/RF (W)	Cell Radius(m)	Bandwidth (MHZ)	Number of Users	Location	Users
Femto Cell	0.01 - 0.25	10 -100	10	1 - 30	Indoor	Homes or small offices
Pico Cell	0.25 - 1	100 -200	20	30 - 100	Indoor/ Outdoor	High-rise buildings, hotels, office buildings or parks
Micro Cell	1 - 10	200 -2000	20 - 40	100 - 2,000	Indoor/ Outdoor	Shopping centres, transportation hubs, city blocks, stadiums, temporary events
Macro Cell	10 - >50	800 - 300000	60 - 75	>2,000	Outdoor	Suburban, city and rural areas

3.2 Analytical System Model

3.2.1 Network Model

A heterogeneous system comprising 4G macro cells and 5G small cells, as shown in Figure 3.2 is adopted as a network model in this work. The cells are randomly scattered under the umbrella coverage of the macro cell to provide high throughput by the line of sight (LOS) communication model. The macro cell is used when a line-of-sight connection could not be established for small cells since there are several obstacles in an urban environment, they can also be used for transmitting control messages in the network. The assumption is that macro and small cells are assigned to different frequency bands, with inter-cell interference occurring only between cells of the same type, in our case we consider interference from small cell types, with their coverage overlapping.

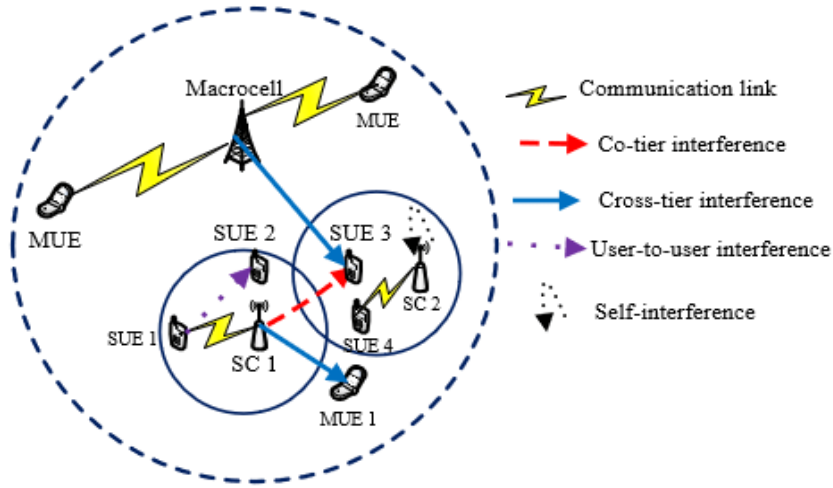


Figure 3.2 Heterogenous 5G Network System Model [44]

3.2.2 Propagation and Channel Model

The signal received by a UE is usually affected by path loss, fading and interference. The link model experienced by the receiver (P_{RX}) can be expressed as in [40] [41]:

$$P_{RX} = P_{TX} + G_T - PL - \sigma \quad (1)$$

where P_{TX} and G_T designate the transmitter power and the total antenna gain respectively, PL denotes the path loss model expanded below, while σ is the shadow fading margin. The standard 3GPP probabilistic LOS propagation is assumed, and the path loss model is stated [30], as follows:

$$PL_{u,k,l} = 20 \log_{10} \left(\frac{4\pi r_0}{\lambda_l} \right) + 20 \log_{10} \left(\frac{d_{u,k}}{d_0} \right) + \chi \quad (2)$$

where, if $l=1$ a small cell is selected, otherwise a macro cell is selected and d_0 and $d_{u,k}$ are the reference distance and the distance between the user u and BS k , correspondingly as shown in Figure 3.3 below, λ_l is the wavelength. χ is a Gaussian random variable with variance σ^2 .

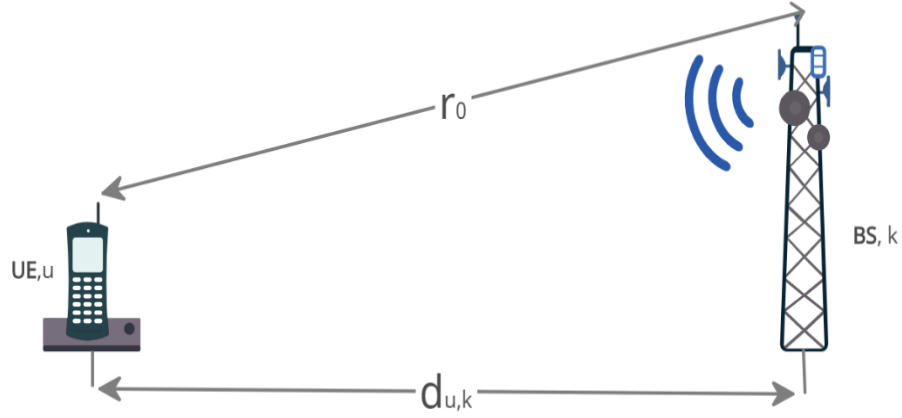


Figure 3.3 System Propagation Model

The UE-Macro cell association incurs Signal to Interference plus Noise Ratio (SINR) which can be modelled as:

$$SINR_{u,k} = \frac{P_{u,k} \times G_{u,k}}{\sum_{k \neq K} P_{u,k} + P_{AWGN}} \quad (3)$$

Where $P_{u,k}$ is the received signal power by UE u from macro base station k , $G_{u,k}$ is the channel gain received by UE u from BS k and P_{AWGN} is the additive white Gaussian noise power. In 5G mmWave band the interference is trivial; it can be ignored and the SNR for the 5G mmWave can be modelled as [30]:

$$SNR_{u,k} = \frac{P_{u,k} \times G_{u,k} \times PL_{u,k}^{-1}}{P_{WGN}} \quad (4)$$

3.2.3 Traffic and Mobility Model

This work models user mobility using the Random Waypoint mobility model. This mobility model describes a mobile network node's movement within a specific system area and is a straightforward stochastic model. A node randomly chooses a destination point (waypoint) within the region and travels there while maintaining a constant speed. It chooses a new destination and speed after a brief pause, moves at a constant speed to this destination, and so on. The system area is covered by a uniform and random distribution of the destination points.

3.3 Problem Formulation

We consider the A3 handover triggering event, in which heterogeneous small cells are laid in an overlapping fashion and a UE located within the coverage area of one cell moves in a straight line into another small cell's coverage region. If the received signal strength (RSS) from the target cell (RSS_t) is greater than the serving cell's signal strength (RSS_s) plus some hysteresis margin, H , a handover will be performed, i.e., when,

$$RSS_t \geq RSS_s + H \quad (5)$$

A handover failure, or via a breakdown of the radio connection, occurs during the handover event if the signal-to-interference-plus-noise ratio (SINR) from the serving small cell falls below a predetermined threshold, Q , for more than one second, i.e.

$$SINR_{sc} \leq Q \quad (6)$$

The proposed handover management scheme is depicted in Figure 3.4 below. The scheme's main aim is to reduce handover failures by leveraging a Semi Markov user mobility algorithm. It consists of three main components, namely, 1) The Mobility Prediction module which aims to predict the future location of a user equipment (UE) and prepare the next base station (BS) 2) An Admission Control Algorithm which determines whether a call should be admitted or dropped/blocked, and 3) A Base Station Selection Algorithm which typically deals with signal strength measurement of base stations utilising handover parameters (Handover Margin and Time to trigger) from the admission control algorithm as well as location estimate from the mobility prediction module.

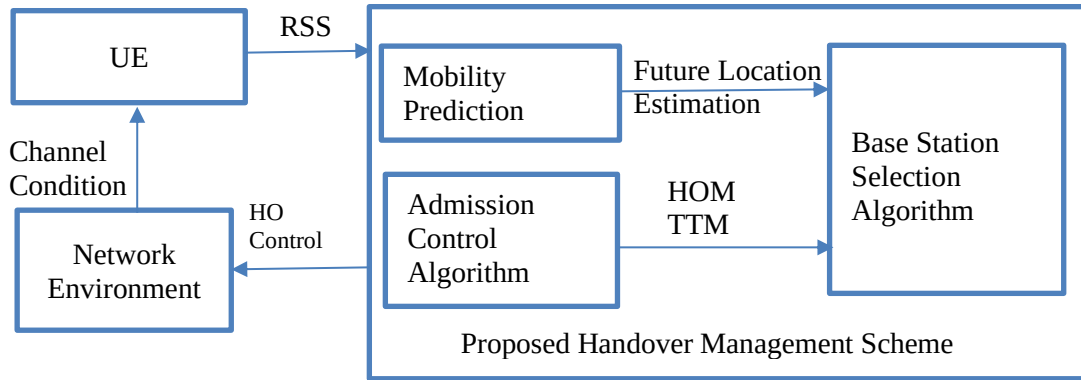


Figure 3.4 Proposed Handover Management Scheme

3.4 Mobility Prediction

3.4.1 Mobility Prediction schemes

When mobility prediction is utilised to foresee a user's long-term movements, the number of handovers, the handover failure rate, and the handover latency may all be significantly decreased (e.g., after changes in transmission points (TPs), the following section of the user's trajectory, and upcoming places). Figure 3.5 below outlines a comprehensive framework of mobility prediction, starting from the bottom are the sources

of data for prediction, followed by the algorithm techniques used in the literature used to interpret such data and output desired information for prediction. Subsequently, performance metrics are used to evaluate the results followed by applications that leverage such metrics at the top to improve the quality of service (QoS) for users [60].

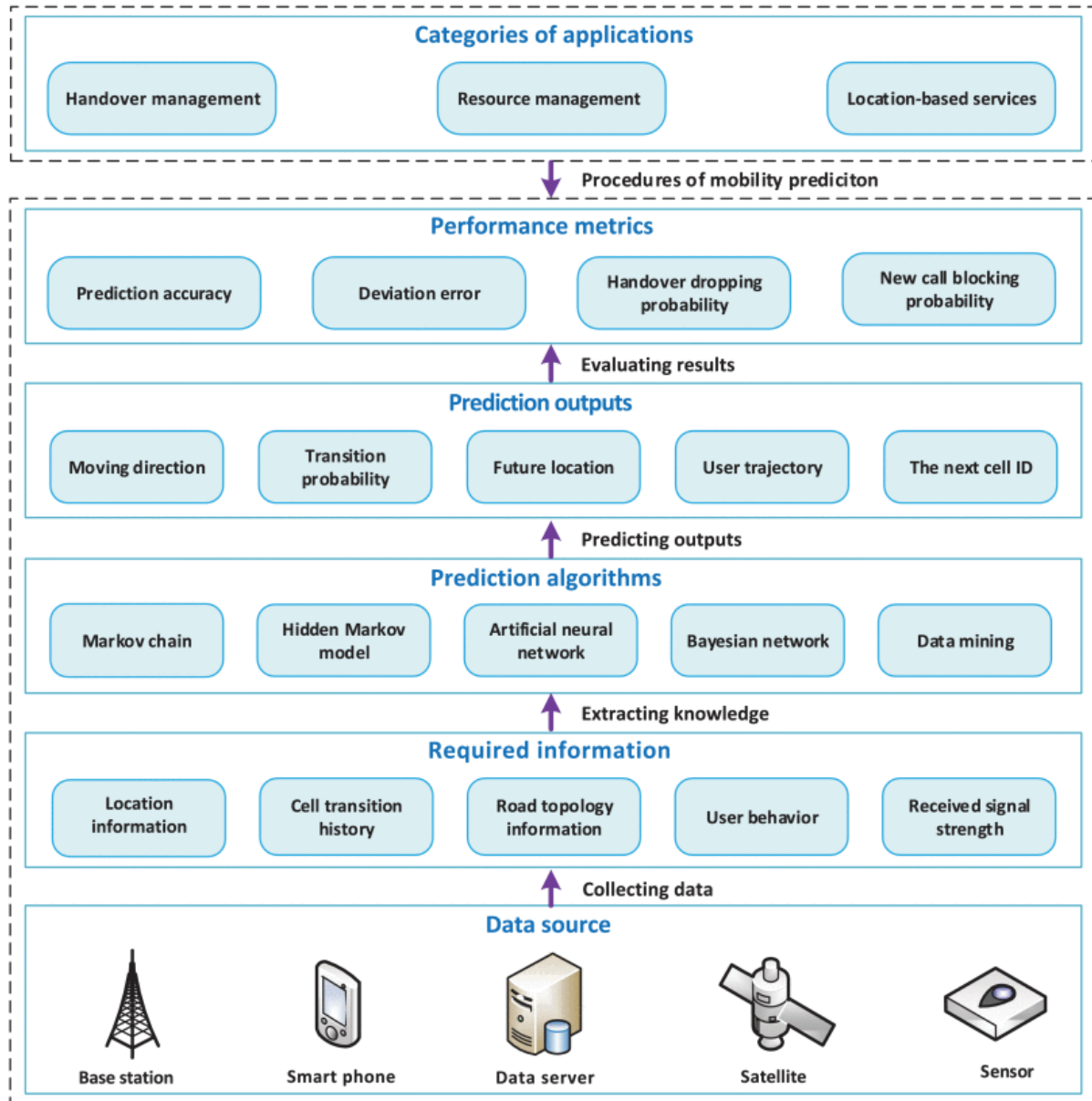


Figure 3.5 An Overall Structure of Mobility Prediction [60]

3.4.1.1 Markov Chain

A Markov chain is an example of a stochastic process having the "memoryless" property. A procedure that goes through a series of random variables is referred to as a "Markov chain." Only the dependency of consecutive periods (as in a "chain") may be used to define serial states according to the Markov property. As a result, it may be used to describe systems where a series of interconnected events occur, with the next event depending on the system's present state or a few previous states. Due to the Markov property, the Markov chain has long been used to anticipate the movability of users in wireless systems.

3.4.1.2 Hidden Markov Model

In conventional Markov model types (such as a Markov chain), the state is easily observable by the viewer, leaving just the state transition probabilities as parameters. The hidden Markov model (HMM) is a twofold random procedure that can only be seen through a distinct collection of stochastic procedures that produce the order of realised symbols. The underlying stochastic process of the HMM is unobservable (hidden). In HMM, there are two distinct sorts of states: observable states and hidden states. Since more data regarding hidden states can be gleaned from visible states, HMM often surpasses the performance of the Markov chain in predicting user mobility.

3.4.1.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) belong to a family models used in machine learning and cognitive science to determine or roughly approximating functions that can depend on several inputs yet are generally unknown. These models were inspired by biological neural networks, specifically the central nervous systems of animals, particularly the brain. The most often used method in artificial neural networks (ANNs), back propagation, is a well-studied algorithm that is renowned for its adaptable and self-organization properties. Furthermore, ANN is renowned for its astounding computational complexity, particularly when there are several hidden layers, which is referred to as a deep neural network and necessitates a significant amount of training time in the tuning of neuron weights.

3.4.1.4 Bayesian Network

A directed acyclic graph is used in the Bayesian network, a probabilistic visual model, to depict a set of random variables and their conditional relationships. In a complex context, it provides considerable advantages in resolving ambiguity and determining the link between various factors. Consequently, Bayesian networks have been widely used in various fields. The Bayesian network, for instance, may be used to determine the likelihood that a user would be present at a particular location given past data on user mobility. Bayesian networks use edges in a directed graph to represent conditional reliance in order to simulate conditional dependence and, by extension, causality. Through the use of factors and these correlations, it is possible to efficiently draw conclusions about the graph's random variables. Exact inference requires a lot of computing in big networks, hence techniques to cut down on computation must be implemented. Additionally, the Bayesian network's development must take into account the cell environment. The cell environment would be more complex as a result of the dense deployment of tiny cells, which would make building a Bayesian network more challenging.

3.4.1.5 Data Mining

Data mining is a computational technique that uses methods from the fields of machine learning, statistics, and database systems to find patterns in vast amounts of data. It is much more accurate to predict the movement trend of the observed user when a big data collection is employed, which includes not just trajectory data but also other user or network-related knowledge.

3.4.2 Proposed Mobility Prediction Scheme

Mobility prediction is a critical and a desirable issue in 5G networks to handle handover management, resources reservation, interference coordination and load balancing. In this work, user mobility is modelled as a Semi-Markov process. A state transition is used to represent the transfer of control from one cell to another. Let, T_n^u be a random variable for the transition's time instant from C_n^u to C_{n+1}^u . The probability user u for transitioning to the j th cell from the i th having stayed for t time is:

$$Q_{i,j}^u(t) = P [C_{n+1}^u = j, T_{n+1}^u - T_n^u \leq t | C_0^u, \dots, C_n^u; T_0^u, \dots, T_n^u] \\ = P[C_{n+1}^u = j, T_{n+1}^u - T_n^u \leq t | C_n^u = i] \quad (7)$$

Where random variable $T_{n+1}^u - T_n^u$ is the sojourn time in a cell which is assumed to be independent from the state transition process $C_{i,j}$. In respect of that, equation (7) reduces to

$$Q_{i,j}^u(t) = P[C_{n+1}^u = j | C_n^u = i] \\ \cdot P [T_{n+1}^u - T_n^u \leq t | C_{n+1}^u = j, C_n^u = i] = \\ \lim_{t \rightarrow \infty} Q_{i,j}^u(t) P [T_{n+1}^u - T_n^u \leq t | C_{n+1}^u = j, C_n^u = i] \quad (8)$$

Here $\lim_{t \rightarrow \infty} Q_{i,j}^u(t)$ is the HO Probability of a user u , transitioning to cell j from cell i and $P [T_{n+1}^u - T_n^u \leq t | C_{n+1}^u = j, C_n^u = i]$ is the distribution of user's sojourn time. Therefore, the probability of u staying in cell i for no more than time t can be expressed as:

$$S_i^u(i) = P [T_{n+1}^u - T_n^u \leq t | C_{n+1}^u = j, C_n^u = i] \\ = \sum_{j=1}^Z Q_{i,j}^u(t) \quad (9)$$

3.5 Admission Control Algorithm

Radio resource management (RRM) maintains the radio link between user equipment (UE) and base stations (BSs) within their coverage area by initiating and executing a handover of the UE from one cell to another while maintaining the session. To begin, the UE sends the measurement report (MR) to the BS on a regular basis, indicating the status of the signal strength power. Second, if the serving BS's average received signal strength falls below the signal strength of the target BS plus handover margin (HOM) for the duration of the time-to-trigger (TTT) interval, the UE triggers handover (HO) and sends the measurement report to its serving BS. Finally, after the serving BS sends the HO information to the target BS, HO is executed. The HOM level between the serving and target BS signal strengths, as well as the TTT interval, prevents the HO ping-pong effect.

Figure 3.6 below shows the flow chart of the proposed handover scheme with the following steps:

Step 1: A heterogeneous network (HetNet) consisting of different types of base stations (BS) and user equipment (UEs) following a Random Waypoint mobility model is initialised.

Step 2: The user mobility prediction framework is used to determine user speed, direction, and handover history.

Step 3: While the user is in motion system/network parameters like path loss and signal strength are continuously measured in anticipation of triggering a handover. That is, handover triggering event A3 (i.e., when Serving BS signal strength is less than Target BS signal strength plus a pre-set threshold) is continuously monitored, and if it is satisfied, the Base Station Selection and Call Admission Control (BS & CAC) algorithm is invoked. The algorithm is expanded in Figure 3.8.

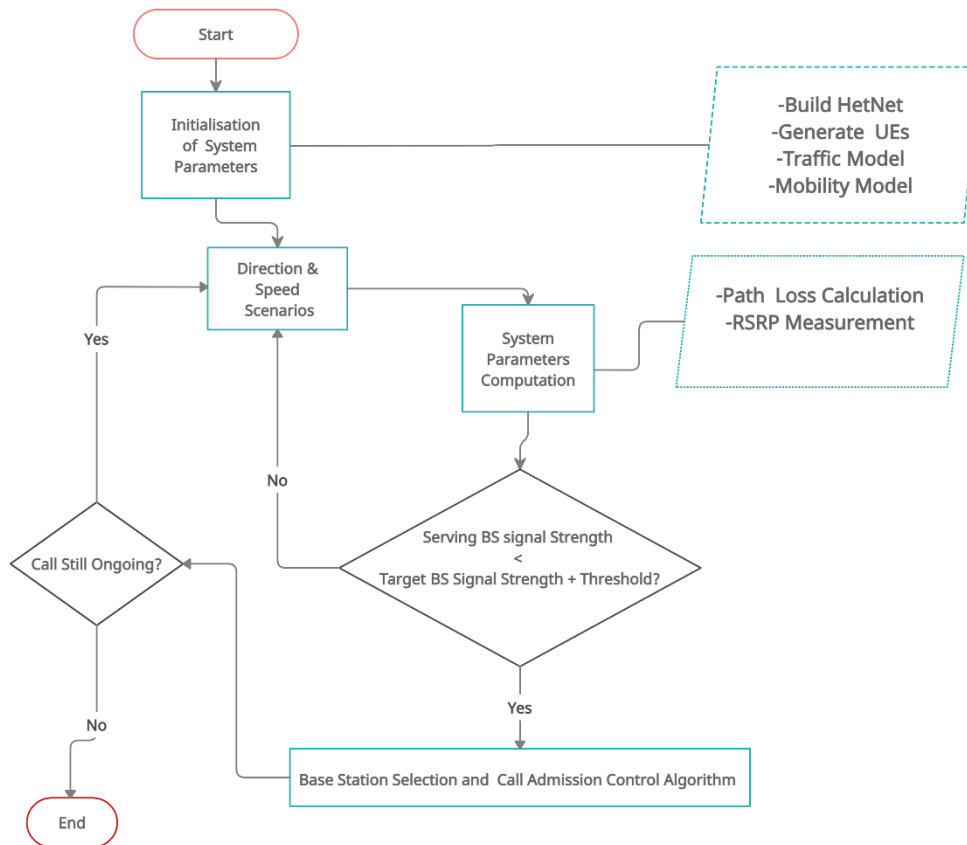


Figure 3.6 Proposed Handover Scheme

3.6 Base Station Selection Algorithm

The current telecommunications industry is being driven by the exponential growth of mobile terminals in cellular communication networks to provide broad information access. This necessitates not only an increase in cellular system capacity but also the expansion of effective call management measures to improve radio resource utilization. In this work, we consider *new calls* and *incoming handover calls*. The calls are handled by the cell's base station, which uses a limited number of radio channels. When a call is initiated or received (new or handover), it is blocked if the base station cannot allocate a free channel that will guarantee the requested level of service. Typically, the dropping of a handover call is exceedingly unwelcome than the blocking of a new call. Figure 3.7 below shows an

illustration of a cell base station admitting and terminating both handover and new calls as considered in this work. Calls that are not completed within this base station are also handed over to other base stations.

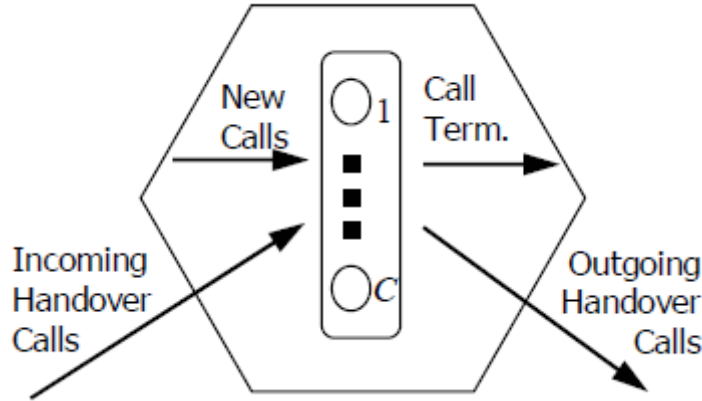


Figure 3.7 Incoming, Terminated and Outgoing Calls in a Cell

The regulation and admission of new and handover calls can be categorised into three different schemes:

- 1) *Complete sharing* - When no priority is given to any type of call, new or handover call.
- 2) *Guard channels* - Handover calls are prioritised over new calls in this scheme by dedicating some channels entirely to them.
- 3) *Handover queueing* - In this scheme handover calls are kept in a queue when no resources are available. When resources are made available, no incoming calls are accepted until the queue is cleared.

The following assumptions are going to be used: The arrival of new and handover calls is governed by Poisson processes with rates λ_1 and λ_2 respectively. An accepted new or handover call occupies one to three radio channels out of the total C channels and remains in the cell for an exponentially distributed time with mean $1/\mu$, where μ is the departure rate of calls. The cell load is thus given by

$$\rho = \frac{\lambda_1 + \lambda_2}{\mu} \quad (10)$$

It must be noted that the departure rate, μ , is the total of calls that are terminated and handover calls, i.e. $\mu = \mu_t + \mu_h$ and $\lambda = \lambda_1 + \lambda_2$. In this work, for simulation and results analysis purposes, we consider a heterogeneous system that consists of two radio access technologies (RATs), RAT-1 and RAT-2. RAT-1 has a capacity (C_1) of 20 basic bandwidth units (bbu), and RAT-2 has a capacity (C_2) of 30bbu, and each of the two RATs supports voice and video calls. A voice call requires 1 bbu, and a video call requires 3bbu. The network has two user groups: Group-A users and Group-B users. Users in Group-A have single-mode terminals that only handle voice and data connections to RAT-1. Users in Group-B are equipped with dual-mode mobile terminals that can establish connections to both RAT-1 and RAT-2.

Now let,

$b_1 \triangleq \text{Bandwidth for Voice call} = 1\text{bbu}$

$b_2 \triangleq \text{Bandwidth for Video call} = 3\text{bbu}$

- $C_1 \triangleq \text{Capacity for RAT} - 1 = 20\text{bbu}$
 $C_2 \triangleq \text{Capacity for RAT} - 2 = 30\text{bbu}$
 $n_1 \triangleq \text{Number of New Voice Calls in RAT} - 1$
 $n_2 \triangleq \text{Number of New Video Calls in RAT} - 1$
 $n_3 \triangleq \text{Number of New Voice Calls in RAT} - 2$
 $n_4 \triangleq \text{Number of New Video Calls in RAT} - 2$
 $h_1 \triangleq \text{Number of Handover Voice Calls in RAT} - 1$
 $h_2 \triangleq \text{Number of Handover Video Calls in RAT} - 1$
 $h_3 \triangleq \text{Number of Handover Voice Calls in RAT} - 2$
 $h_4 \triangleq \text{Number of Handover Video Calls in RAT} - 2$

Admissible States:

The heterogeneous cellular network above can only accept incoming new, and handover calls if:

$$b_1 n_1 + b_1 h_1 + b_2 h_2 + b_2 n_2 \leq C_1 \wedge b_1 n_3 + b_1 h_3 + b_2 n_4 + b_2 h_4 \leq C_2$$

Blocking States (Group-A):

Voice Calls:

$$b_1 + b_1 n_1 + b_1 h_1 + b_2 n_2 + b_2 h_2 > C_1$$

Video Calls:

$$b_2 + b_1 n_1 + b_1 h_1 + b_2 n_2 + b_2 h_2 > C_1$$

Blocking States (Group-B):

Voice Calls:

$$b_1 + b_1 n_1 + b_1 h_1 + b_2 n_2 + b_2 h_2 > C_1 \wedge b_1 + b_1 n_3 + b_1 h_3 + b_2 n_4 + b_2 h_4 > C_2$$

Video Calls:

$$b_2 + b_1 n_1 + b_1 h_1 + b_2 n_2 + b_2 h_2 > C_1 \wedge b_2 + b_1 n_3 + b_1 h_3 + b_2 n_4 + b_2 h_4 > C_2$$

Overall blocking state for the Network (Voice +Video) Calls:

$$b_1 + b_2 + b_1 n_1 + b_1 h_1 + b_2 n_2 + b_2 h_2 > C_1 \wedge b_1 + b_2 + b_1 n_3 + b_1 h_3 + b_2 n_4 + b_2 h_4 > C_2$$

Consequently, the blocking probabilities for new calls (voice and video), P_n , and handover calls (voice and video), P_h , can be obtained using the Erlang formula as follows (only for RAT-1 for illustration purpose):

$$P_n = \frac{\left(\frac{(\rho_{n_1})^{n_1} (\rho_{n_2})^{n_2}}{n_1! n_2!} \right)}{\sum_{s \in s_a} \left(\frac{(\rho_{n_1})^{n_1} (\rho_{n_2})^{n_2}}{n_1! n_2!} \right)} \quad (11)$$

, where s_a denotes set of Admissible states, n_1 is the quantity of voice calls and n_2 is the number of new video calls.

The Normalisation Factor G is given by:

$$G = \sum_{s \in s_a} \left(\frac{(\rho_{n_1})^{n_1} (\rho_{n_2})^{n_2}}{n_1! n_2!} \right) \quad (12)$$

Hence, the probability of blocking handover calls will be given by:

$$P_h = \frac{\left(\frac{(\rho_{h_1})^{h_1} (\rho_{h_2})^{h_2}}{h_1! h_2!} \right)}{\sum_{s \in s_a} \left(\frac{(\rho_{h_1})^{h_1} (\rho_{h_2})^{h_2}}{h_1! h_2!} \right)} \quad (13)$$

The above formulations are incorporated into the base station selection algorithm in Figure 3.8 below. Depending on the capacity of a RAT, and the type of call (voice or video), a RAT is in an admissible state if its capacity has not been reached due to a current number of calls that are in session, and while in that state the call is admitted. When a RAT's capacity cannot accommodate the bandwidth requirements of a dual-mode user terminal, the additional bandwidth is sourced from a neighbouring RAT, but for a single mode device, the call is dropped or blocked for handover or new calls respectively. After a call has reached its duration, it is terminated and the resources it occupied are released. For calls that are still ongoing, the speed, direction and the handover triggering event are monitored as in Figure 3.6 in preparation for a handover.

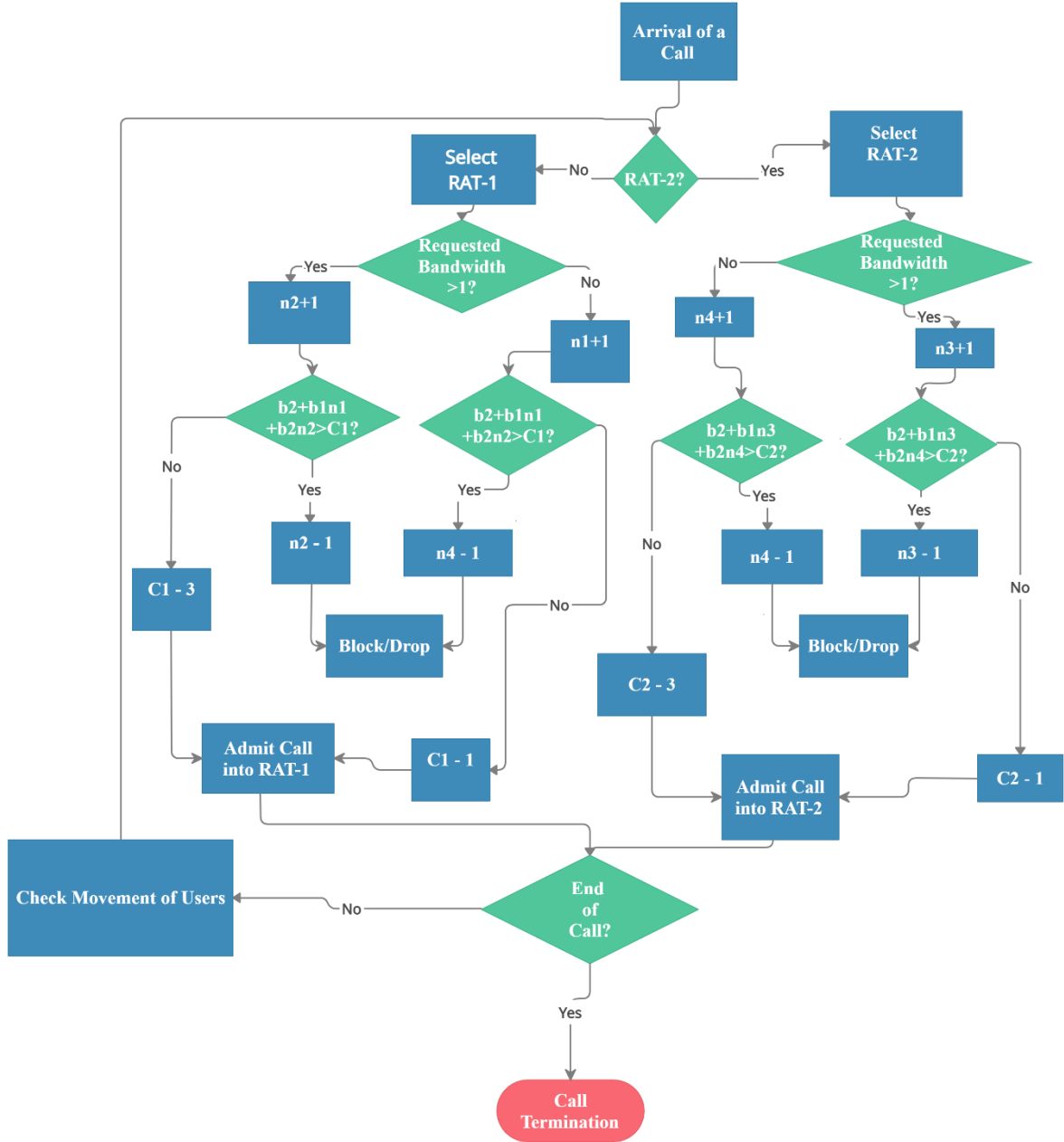


Figure 3.8 Base Station Selection and Call Admission Control Algorithm

In the literature, several queuing models that use the First in First out (FIFO) discipline have been utilized, which include M/M/1 model, M/M/c model, M/M/c/K model and M/M/c/∞/N model. In the 5G space, most works propose and utilise different variants of the M/M/c model for 5G to cater for efficient and scalable traffic evidenced in 5G networks. In this work, the classical M/M/c model is used as it is sufficient to meet the objectives of a handover scheme that aims to reduce handover failure. The choice was influenced by its simplicity and that the main objective of the study was to cater for load balancing and the massive traffic in 5G which is also susceptible to fluctuations.

3.7 Chapter Summary

In this chapter, 5G heterogeneous networks are introduced and briefly explained. Subsequently an analytical model is proposed which incorporates assumptions for network, propagation and channel, and traffic and mobility models. Thereafter, problem formulation follows which sets the tone for the proposition of a mobility prediction scheme for fulfilling the objectives of this work. In conjunction with the proposed scheme, two algorithms, namely the Admission Control (ACA) and a Base Station Selection Algorithm (BSSA) are proposed. The algorithms are developed in MATLAB and a simulation environment created and tested, whose results and analysis are presented in the following chapter.

Chapter 4

Simulation and Results

This chapter outlines the simulation setup and environment using MATLAB. For testing and analysing results, a handover from one small cell to another is considered.

4.1 The Simulation Environment

The simulations are performed in MATLAB. A communication environment with a size of 30 m x 30 m. The testing environment is built on the standards outlined in Report ITU-R M.2412 [58] and Report ITU-R M.2135-1 [59] for evaluating 5G radio technologies as depicted in Figure 4.1 below. The two base stations (BS) are denoted by the red balloon structures while an instance of user equipment (UE) is illustrated by the blue balloon structure which moves in a horizontal trajectory (shown with a thick white stripe). The green and red lines depict the link establishment and a link not established respectively. Dual connectivity is established at a point in time when a green link is established simultaneously for both BSs and once the UE moves out of the range of either of them, the data packets are transferred/handed over to another BS. The movement of the UE is generated by utilising the random waypoint function from MATLAB Central File Exchange [62]. There are 20 UEs randomly moving at a speed randomly selected in a range of 0.2 to 2.2 m/s; such speed resembles the average walking speed of a human. The A3 event Received Signal Reference Power (RSRP) based handover scheme and a Base Station Selection and Call Admission Control (BSCAC) algorithm are used to complete the handover process. The assumption of the mobility model which is random waypoint is utilised in conjunction with call arrival and sojourn times, which Poisson process and exponential distribution respectively. Here the idea is that the human individual carrying a UE is modelled by random waypoint movement but those users who will initiate calls and the ongoing calls that need to be handed over are modelled as Poisson processes and their sojourn times as being exponentially distributed. Table 4.1 shows the key simulation parameters.

Table 4.1 Simulation Parameters

Simulation Area	30 x 30 m
Traffic Model	Voice and Video Calls
Mobility Model	Random Waypoint
BS Tx Power	10 dBm
Antenna	Isotropic
BS Antenna Height	10 m
BS1 Location	42.3763°, -71.0611°
BS2 Location	42.3467°, -71.0972°
Tx Frequency	2.5 MHz
Bandwidth	20 MHz and 30 MHz
Number of UE	20

UE Receiver Sensitivity	-90 dBm
UE Antenna Height	1 m
UEs Speed Interval	[0.2 2.2] m/s
Handover Event	A3
A3 Threshold	3 dB
TTT/T310	1 s
Path Loss Model	Free Space
Simulation Time	500 s

4.2 Evaluation Metrics

Handover algorithms and schemes are compared using several parameters. These are a number of handovers, probability of handover failure, throughput, and latency. Handover Delay: It is the period of time between the start and finish of a handover process. It is especially important in time-sensitive applications. Number of Handovers: Frequent and unnecessary handovers are not desirable since they result in a wastage of resources and power consumption, hence they need to be minimised. Probability of Handover Failure: Handover failure ensues once a handover process is started but the mobile user goes out of the selected network's coverage region, or the selected network lacks enough radio resources to carry out the handover process. According to [47], a handover failure (HOF) can be declared at a UE if any of the three criteria listed below is met:

- 1) A radio link failure (RLF) happens during the interval between fulfilling a measurement triggering event (in this case the A3 event) condition and receiving the handover command.

- 2) When a handover command is sent, the failure of the physical down link control channel (PDCCH) is detected.

- 3) When a handover complete message is sent, the signal to interference plus noise (SINR) of the UE wideband SINR is less than a threshold.

Throughput: The data rate that a base station on the network provides to a mobile user. Handover is generally preferred on networks with higher throughput. For analysis and results, handover from one small cell to another with two different radio access technologies is used as depicted from Figure 4.1 below. The cells coverage areas are overlapping.

Handovers are generally classified as early, delayed, and ping-pong handovers. It must be noted that delayed and ping-pong handovers are the most likely to cause call drops. In the case of a delayed handover, the target cell may not have the UE context and as a result, the handover call will be dropped. Normally, the UE reports the A3 measurement report of the neighbouring cell before the call drop, and the gNodeB receives the measurement report and delivers the handover command. However, the UE cannot receive the handover command. In this case, if the target cell has the UE context and the RRC connection can be successfully re-established, the call drop does not occur. To solve call drops caused by delayed handovers, a mobility prediction algorithm can be implemented so that handovers can be performed in advance.

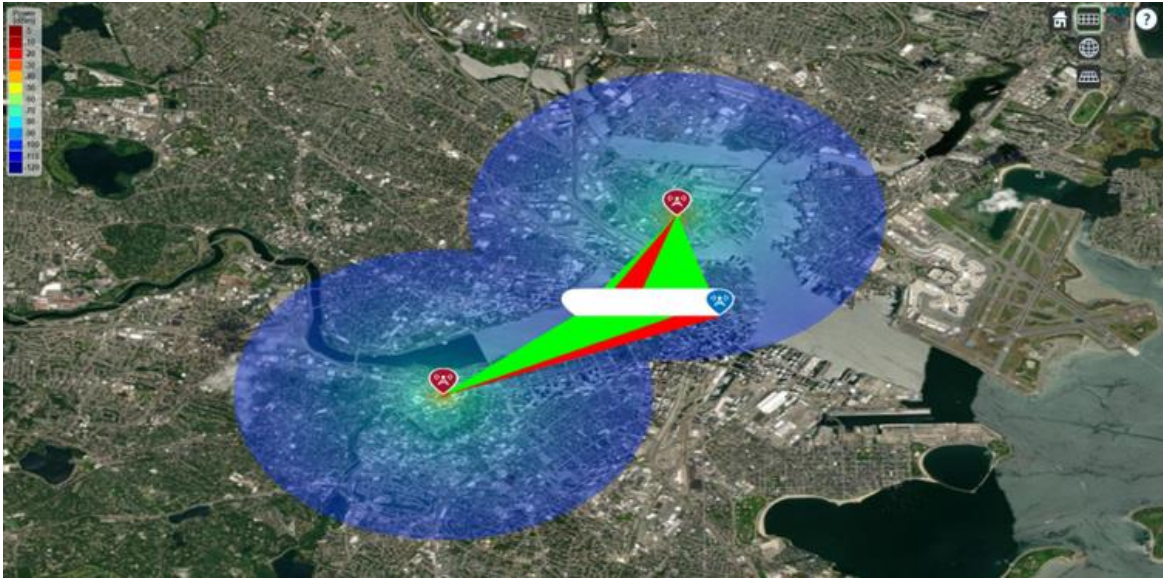


Figure 4.1 Handover from RAT1 to RAT2

Implementing mobility prediction, according to the authors of [51], is an effective method for managing mobility and sustaining users' uninterrupted connections, and its applications are in the management of handovers, resource allocation, and location-based services. The consequence of handover is a typical challenge for wireless systems. Once a user moves out of the current cell coverage area, but it is not able to join a new cell smoothly, unacceptable delays or even call dropping issues may occur, requiring appropriate handover management. Moreover, the implementation of a reactive resource reservation strategy is the only way to guarantee uninterrupted services devoid of the need to reserve a massive amount of resources throughout the whole network [52]. This approach can pre-reserve a specified amount of bandwidth in coverage areas that users are likely to visit, avoiding reserving bandwidth throughout the whole system and thus saving a lot of unnecessary radio resources. Likewise, if the network can accurately forecast users' future trajectories, other location-based aspects that are extremely linked to users can be distributed, and the network will be able to predict traffic conditions as well as design more efficient strategies.

In the existing literature, various methods and performance metrics are utilised to predict user mobility. Herein, the key focus is on the metrics that are used to quantify the performance of the proposed algorithms and schemes. Metrics are classified into two types: observable metrics and application metrics. Observable metrics, such as prediction accuracy and deviation error, can be obtained directly from the prediction result, whereas application metrics, such as handover dropping probability and new call blocking probability, are used to describe the performance of the prediction result's application. Application metrics are utilised herein for measuring the performance of the proposed scheme. The handover dropping probability is calculated as the ratio of dropped handovers to total handovers. An indirect assessment of the prediction schemes is made since the forecast result is used for handover management. In resource reservation schemes, the call blocking probability is invariably increased in order to meet the handover dropping probability aim.

4.3 Simulation Results and Evaluation

4.3.1 Result of changing the call arrival rate on the new call blocking probability and handover call dropping probability

The effect of increasing the influx of incoming calls (from 1 call per minute to 10 calls per minute) on call blocking and dropping probabilities of new and handover calls for each group of users, given that the bandwidth of voice calls is 1bbu, the bandwidth of video calls is 3 bbu, the arrival rate of both the new and handover calls as 2 calls per minute, the departure rate of handover calls as 0.5 calls/minute, and the departure rate of new call as 1 call per minute was evaluated using MATLAB and the performance is shown in Figure 4.2 below. It shows that call blocking and dropping probabilities of new and handover calls for each group of users increase with the call arrival rate increment, but the probability of dropping handover calls is always lower than that of blocking new calls because handover calls are given priority over new calls which are just initiated. More precisely, when the arrival rate of the calls is increased from 1 call per minute to 10 calls per minute, the probability of dropping handover calls steadily increase, this is due to the fact that the network is experiencing a high number of ongoing calls than the ones that are leaving the system. It should also be noted that in the same breath, the blocking of new calls is much higher than the dropping of handover calls, this is so, because with handover calls, their mobility is predicted prior to execution, and hence radio resources booked for them. Another important observation is that group B users calls have a lower probability of being blocked/dropped due to the fact that the user terminals are able to connect to two base stations simultaneously when their bandwidth requirements cannot be met by a single base station.

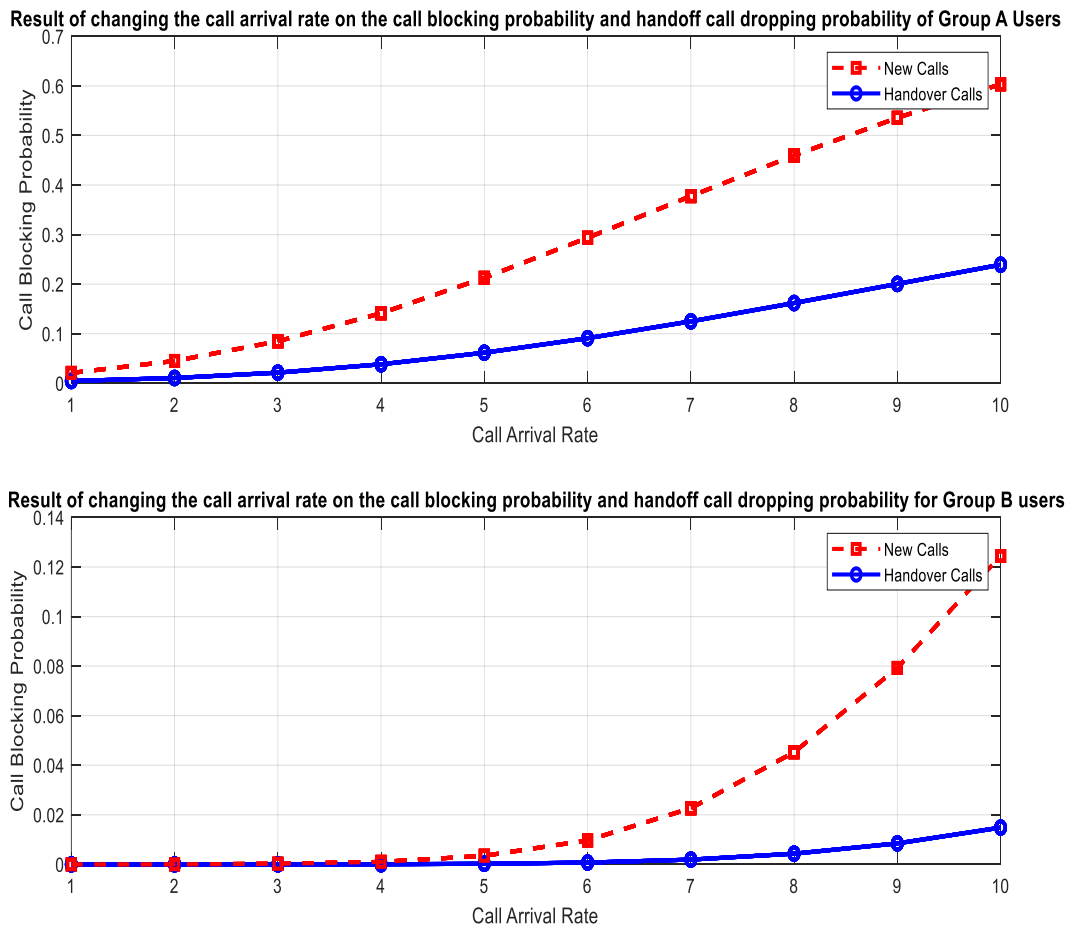


Figure 4.2 Impact of changing the call arrival rate on the call blocking probability and handover call dropping probability.

4.3.2 Result of increasing network capacity on the new call blocking probability and handover call dropping probability

The effect of increasing network capacity (by 10% to 100%) on call blocking and dropping probabilities of new and handover calls for each group of users, given that the bandwidth of voice calls is 1 bbu, the bandwidth of video calls is 3 bbu, the arrival rate of new calls is 2 calls per minute, the arrival rate of handover calls is 2 calls per minute, the departure rate of handover calls is 0.5 calls per minute, and the departure rate of new calls is 1 call per minute was evaluated using MATLAB and the performance is shown in Figure 4.3 below. It shows that call blocking and dropping probabilities of new and handover calls for each group of users decreases to 0 as the network capacity is increased, and the probability of dropping handover calls is always lower than that of blocking new calls because handover calls are given priority over new calls which are just initiated. Also, the probability of blocking/dropping calls from Group B users is significantly lower than that of group A users due to using multimode devices that can connect to both RATs if the quality of service from one is not sufficient. Another aspect observable is that, since the arrival of calls is kept constant and occupies the same radio resources while the available resources in the system are gradually increased, this creates extra resources that are idle hence this results in no calls being dropped or blocked. It can also be noted that since handover calls are predicted a priori, their dropping probability is always low and once the change is effected,

the probability drops almost immediately to zero, this phenomenon is even more apparent in group B users since they have an alternative of connecting to two base stations simultaneously.

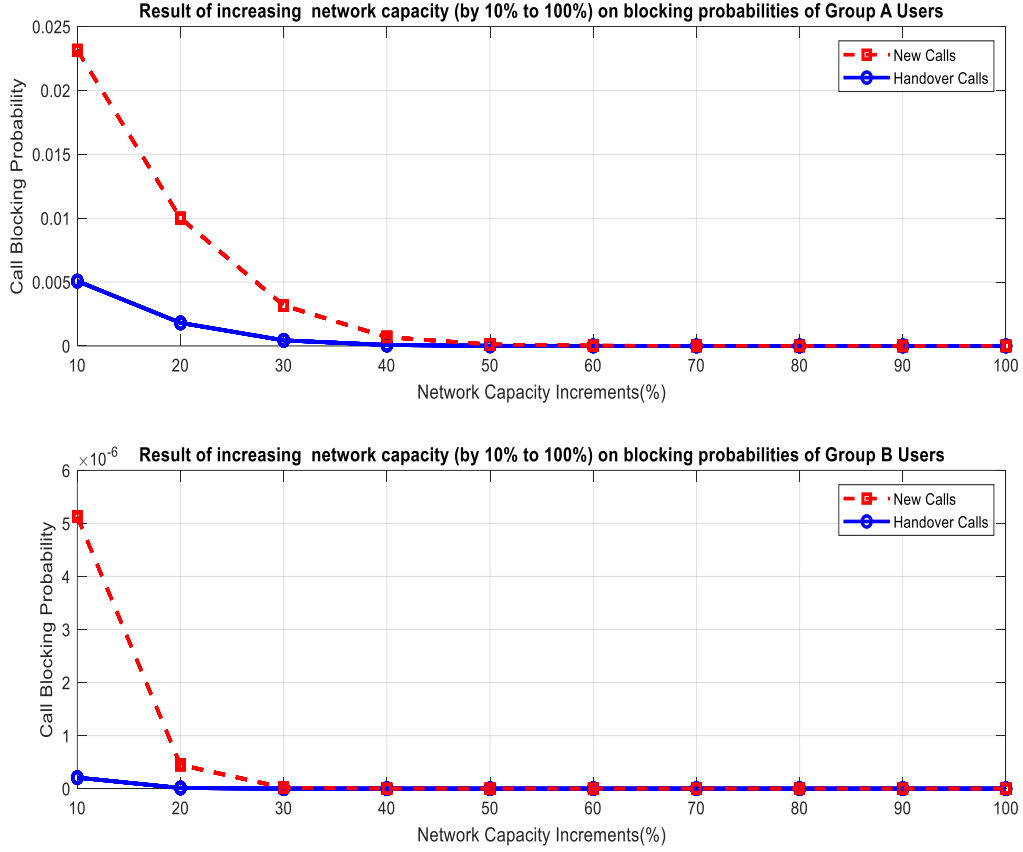


Figure 4.3 Impact of increasing network capacity on the new call blocking probability and handover call dropping probability.

4.3.3 Result of increasing departure rate of handover calls on the new call blocking probability and handover call dropping probability.

The effect of increasing the calls departure rate of handover calls (from 0.5 call/minute to 3 calls/minute) on call blocking and dropping probabilities of new and handover calls for each group of users, given that the bandwidth of voice calls is 1bbu, the bandwidth of video calls is 3 bbu, the arrival rate of new calls is 2 calls per minute, and the departure rate of the new call is 1 call per minute was evaluated using MATLAB and the performance is shown in Figure 4.4 below. It shows that call blocking and dropping probabilities of new and handover calls for each group of users decreases as the departure rate of handover calls is increased, and the probability of dropping handover calls is always lower than that of blocking new calls because handover calls are given priority over new calls which are just initiated. Also, the probability of blocking calls from Group B users is significantly lower than that of group A users due to using multimode devices that can connect to both RATs if the quality of service from one is not sufficient. In this scenario, it is evident that the increase

in handover calls departure rate has effectively translated into a reduction of handover call failure by virtue of decreasing the probability of handover calls dropping.

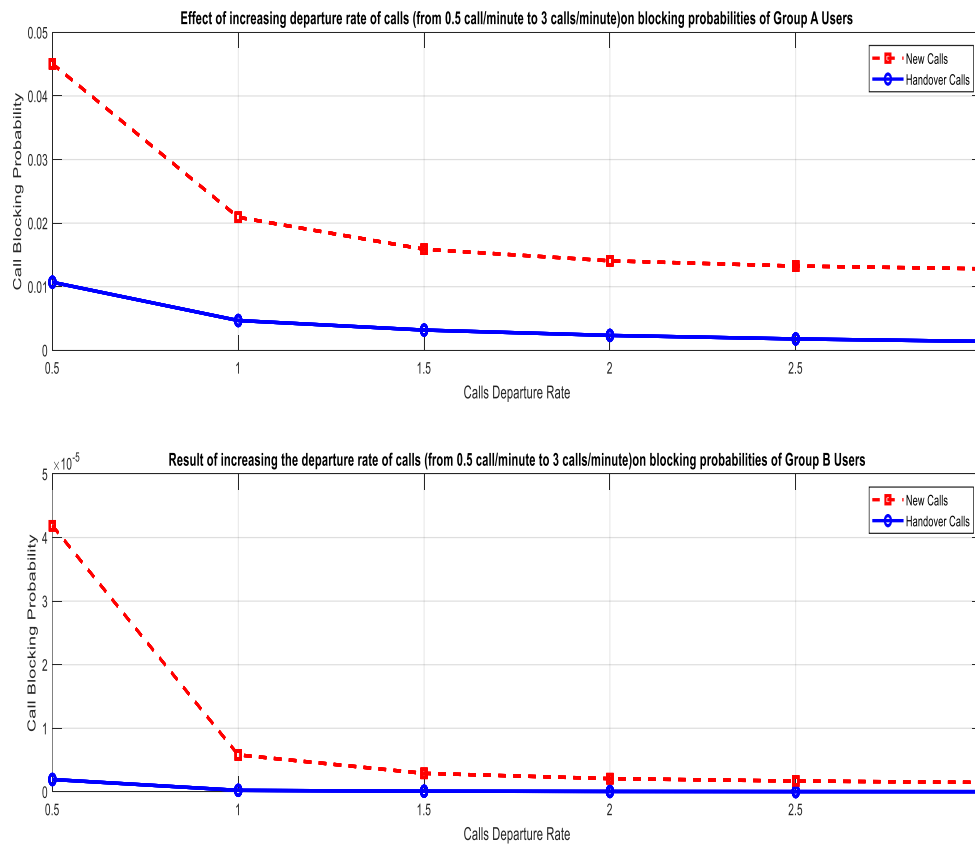


Figure 4.4 Impact of increasing departure rate on the new call blocking probability and handover call dropping probability.

4.3.4 Result of changing the basic bandwidth unit of video calls on the new call blocking probability and handover call dropping probability.

The effect of varying the basic bandwidth unit of video calls (from 1bbu to 6bbu) on call blocking and dropping probabilities of new and handover calls for each group of users, given that the bandwidth of voice calls is 1bbu, the arrival rate of handover calls is 2 calls per minute, the arrival rate of new calls is 2 calls per minute, the departure rate of handover calls is 0.5 calls/minute, and the departure rate of new calls is 1 call per minute was evaluated using MATLAB and the performance is shown in Figure 4.5 below. It shows that call blocking and dropping probabilities of new and handover calls for each group of users increases with the increase of video call bandwidth, but the probability of dropping handover calls is significantly lower than that of blocking new calls in this case. Also, the probability of blocking and dropping calls from Group B users is lower than that of Group A users due to using multimode devices that can connect to both RATs if the quality of service from one is not sufficient. Here the evaluation metric of increasing the bandwidth occupancy of video calls means the capacity of a cell will be quickly reached with fewer calls and hence it is expected that new and handover calls to this cell will be frequently blocked or dropped if there are video calls ongoing. From the graphs, it is evident that since the blocking and

dropping probabilities are increasing but as expected, group B users experience fewer droppings and blockings as compared to group A users. Also, the rate of dropping handover calls is lower than the blocking of new calls in both instances. This also solidifies the fact that the mobility prediction of handover calls is effective.

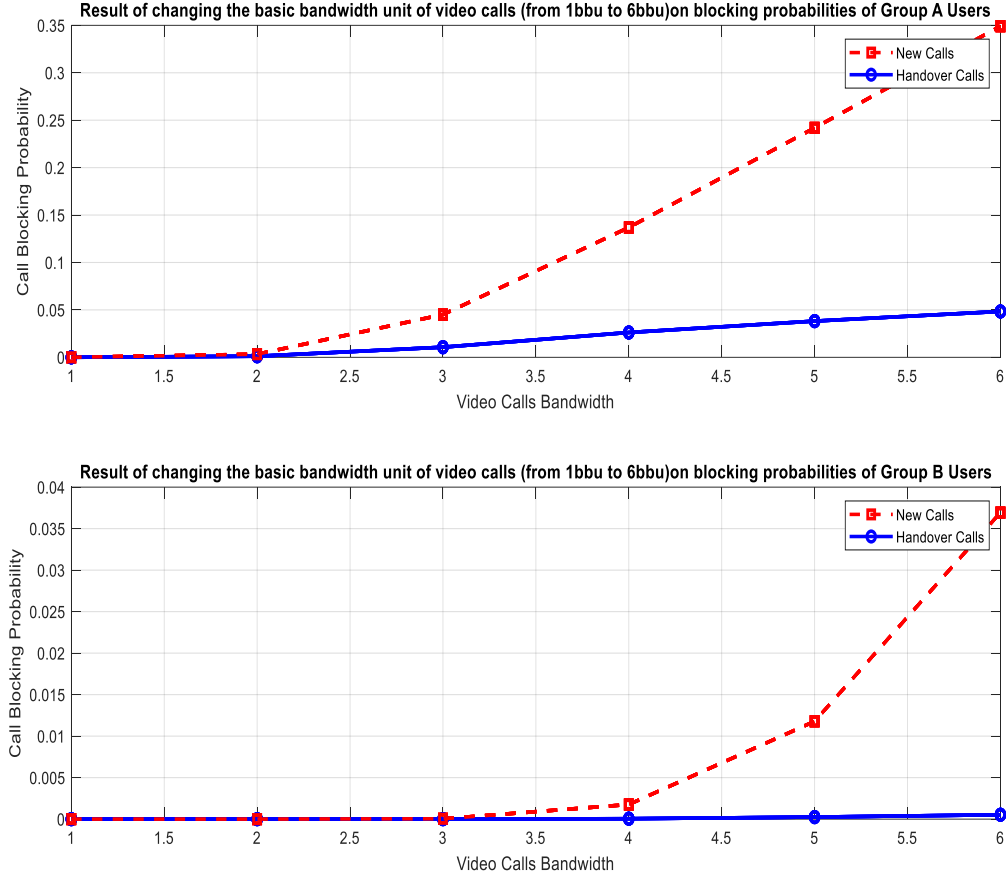


Figure 4.5 Impact of changing the basic bandwidth unit of video calls the new call blocking probability and handover call dropping probability.

In this work, we have leveraged a semi-Markov process prediction scheme to increase the accuracy of the forecast and further improve network performance in terms of user latency, blocking, and handover call dropping probability. As evidenced in the simulation results, the prediction of the next location of the user equipment yielded lower handover call dropping as compared to new calls blocking as they were not predicted and hence resources not reserved for them. The impact of the prediction technique can adequately be evaluated in the test scenarios where the network is bombarded with an unusually high influx of calls and increasing the bandwidth of video calls. In both scenarios, as expected, the call dropping and blocking increased, but because of the prediction for handover calls, their blocking and dropping is still significantly lower than the call blocking of new calls. In our approach, prediction error was not evaluated as it was beyond our scope of work, nonetheless, it produced satisfactory results in our area of focus.

4.3.5 Effect of including mobility prediction versus a system with no mobility prediction

The literature on 5G Mobility/handover Management shows that conventional handover techniques require a lot of time from initiation step to network selection and execution step as compared to systems that utilise mobility prediction. In reviewed works, the advantages were duly noted in [26] and [31] amongst others. In [26], improved on their previous work by proposing a velocity-based self-optimisation algorithm to regulate the HO control parameters in 4G/5G HetNets and address the issue of Mobility Robustness Optimization (MRO) by concentrating on two performance metrics: radio link failure (RLF) and handover ping pong (HOPP). Upon evaluation, the scheme attained a low call drop rate and reduced HO delay as well as a reduction in HO interruption time in comparison with their previous work. Whereas, in [31] the authors present a novel prediction-based conditional handover (PCHO) plan using deep-learning technology to forecast the future BS and overcome the shortcomings of baseline CHO in order to reduce Handover Failure (HOF). The CHO scheme yielded a reduction in signalling overheads as compared to the baseline CHO.

4.4 Chapter Summary

In this chapter a simulation environment was created in MATLAB to validate the scheme proposed in chapter 3. The simulation environment consists of 20 UEs randomly moving at a speed randomly selected in a range of 0.2 to 2.2 m/s; such speed resembles the average walking speed of a human. The A3 event Received Signal Reference Power (RSRP) based handover scheme and a Base Station Selection and Call Admission Control (BSCAC) algorithm are used to complete the handover process. To evaluate the proposed scheme, different network parameters were changed to gauge the behaviour of the scheme. Application metrics are utilised herein for measuring the performance of the proposed scheme. The handover dropping probability is calculated as the ratio of dropped handovers to total handovers. An indirect assessment of the prediction schemes is made since the forecast result is used for handover management. In resource reservation schemes, the call blocking probability is invariably increased to meet the handover dropping probability aim. Application metrics, such as handover dropping probability and new call blocking probability, are used to describe the performance of the prediction result's application. The simulation results showed that the prediction of the next location of the user equipment yielded lower handover call dropping as compared to new calls blocking as they were not predicted and hence resources not reserved for them.

Chapter 5

Conclusion and Recommendations

This chapter contains a thorough discussion of the dissertation's research material. This discussion also includes final remarks on proposed handover algorithms. This chapter concludes with a discussion of future work improvements.

Despite the fact that the analytical model and simulation framework make several simplifying assumptions, our findings provide several useful insights that are supported by simulation results.

5.1 Conclusion

In telecommunications networks when a handover failure occurs, as the user equipment (UE) session will not have been completed, a reconnection is initiated. The reasons that a handover failure may occur, among them, are handover delay, interference, and lack of adequate radio resources in the target cell to handle the demands of the handover session. Any sort of handover typically involves three primary phases: measurement and initiation, decision, and execution. The majority of extant studies classify handover decision systems into five categories: RSS-based schemes, QoS-based schemes, Decision Function-based schemes, Network Intelligence-based schemes, and Context-based schemes. Because of its practicality and scalability, RSS-based handover schemes are being used widely around the world. The user is "selfish" because it chooses the "best" BS according to its RSS without consideration for other users, which lowers the network's overall performance.

This study focuses on three objectives, namely: Examine the current state of the art in 5G mobility management and provide a critical examination of current proposals for alternative techniques in the context of 5G heterogeneous networks, Explore the gaps in 5G handover schemes that aim to reduce handover failure and design and model a handover scheme for heterogeneous 5G networks that aims to improve Handover Failure (HOF). While considering the first two objectives, a critical review of related works was conducted and in addition, an evaluation was conducted for those studies. The qualitative evaluation of the works resulted in an exploration of the gaps that were evident. The gaps formed a foundation for the proposed scheme which seeks to reduce handover failures in 5G heterogeneous networks. The proposed handover scheme includes a mobility prediction algorithm.

In particular, the proposed scheme focuses on the situation where a handover failure occurs as a result of a handover call being dropped or blocked due to the target cell capacity being reached or insufficient resources being available to accommodate the handover call fully. To that end, a simulation environment consisting of small cells with different radio access technologies was set up in MATLAB to evaluate a handover scheme to reduce handover failure by reducing the number of dropped handover requests using a mobility prediction algorithm. The scheme includes a location management mechanism for tracking the UE's current location, as well as a resource management mechanism for controlling the

admission of handover calls and managing radio resource reservations to ensure the same quality of service as the originating base station. The scheme aims to reduce handover failure by incorporating a mobility prediction algorithm and reserve resources by prioritising handover calls over new calls. The prediction framework is based on a semi-Markov renewal model leveraging spatiotemporal properties of a UE in respect of nearby base stations. This allows network resources to be reserved in future identified cells for seamless handover and traffic forecasting. The mobility prediction is based on the assumption that, despite the apparent randomness of individual trajectories, mobile user mobility patterns are, to some extent, predictable. Simulation results show that the prediction algorithm substantially helps lower the handover call dropping probability compared to the new call blocking probability, which does not use the prediction algorithm hence resources are not reserved before their initiation.

5.2 Recommendations

In this study, a semi-Markov prediction algorithm was used which can predict the next base station (BS) with higher accuracy. It must be noted that even though the leveraged prediction algorithm has a high prediction accuracy due to the fact that more than one BS can be prepared to accommodate handover calls, it turns out to be not resource effective. When more than one BS is prepared that means new calls will be dropped even though the radio resources have not been fully utilised. This case is even more extravagant when only one BS will be utilised after the execution of a handover yet more new calls would have been dropped as result, and this factor amplifies in case of incorrect predictions. Therefore, it is pertinent for future work to improve the robustness of the prediction algorithm, it will be interesting to include obstacles and incorporate the use of artificial intelligence (AI) leveraging the reinforcement learning (RL) paradigm in combination with the semi-Markov algorithm to compensate for the shortages of the semi-Markov algorithm since with obstacles the learning nature in AI will keep a record of where an optimal selection of a BS was not reached. The classical M/M/c queuing model was assumed in this work, even though for the scope of this work it was sufficient, it would be befitting the cater for changing nature of the 5G traffic and utilise the M/M/c/ ∞ which will cater for efficiency and scalability.

References

- [1]. Paudel, Prabesh, and Abhi Bhattarai. "5G telecommunication technology: history, overview, requirements and use case scenario in context of Nepal." In *Conference: IT4D*. 2018.
- [2]. Infrastructure Association. "5G Vision-The 5G infrastructure public private partnership; the next generation of communication networks and services." *Available on www. 5g-ppp.eu/roadmaps*. 2015.
- [3]. Paul, C. E. "Telephones aboard Metroliner." *Bell Laboratories Record* 47, no. 3 (1969): 76.
- [4]. Amos Jr, Edward Joel. "Mobile communication system." U.S. Patent 3,663,762 issued May 16, 1972.
- [5]. Akyildiz, Ian F., Jiang Xie, and Shantidev Mohanty. "A survey of mobility management in next-generation all-IP-based wireless systems." *IEEE Wireless communications* 11, no. 4 (2004): 16-28.
- [6]. Akyildiz, Ian F., Janise McNair, Joseph SM Ho, Huseyin Uzunalioglu, and Wenye Wang. "Mobility management in next-generation wireless systems." *Proceedings of the IEEE* 87, no. 8 (1999): 1347-1384.
- [7]. Xie, Jiang, and Shantidev Mohanty. "Mobility management in wireless systems." *Telecommunication Systems and Technologies-Volume II* 5 (2009): 162.
- [8]. Tabassum, Hina, Mohammad Salehi, and Ekram Hossain. "Mobility-aware analysis of 5G and B5G cellular networks: A tutorial." *arXiv preprint arXiv:1805.02719* (2018).
- [9]. Lin, Yi-Bing, Yieh-Ran Haung, Yuan-Kai Chen, and Imrich Chlamtac. "Mobility management: from GPRS to UMTS." *Wireless Communications and Mobile Computing* 1, no. 4 (2001): 339-359.
- [10]. Alsaedy, Alaa AR, and Edwin KP Chong. "A review of mobility management entity in LTE networks: Power consumption and signalling overhead." *International Journal of Network Management* 30, no. 1 (2020): e2088.
- [11]. 3GPP TS 136 300 - V9.4.0. "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description" ETSI TS 136 300 V9.4.0 Release 9 (2010-07)
- [12]. 3GPP TS 136 304 - V13.0.0. "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode " ETSI TS 136 304 V13.0.0 Release 13(2016-02).
- [13]. Access, Evolved Universal Terrestrial Radio. "Study on Small Cell Enhancements for E-UTRA and E-UTRAN—Higher layer aspects (Release 12),". (2013).
- [14]. Mobile subscriptions Q2 2020 "Ericsson Mobility Report", 2020.
- [15]. Agubor, C. K., E. E. Atimati, and A. O. Akande. "5G Wireless Network Mobility Management and Security Issues: An Overview of Existing Technologies." *Futo Journal Series* 4, no. 2 (2018): 84-93.
- [16]. Patel, Chirag, Mehmet Yavuz, and Sanjiv Nanda. "Femtocells [industry perspectives]." *IEEE Wireless Communications* 17, no. 5 (2010): 6-7.
- [17]. Sarnecki, Joseph, C. Vinodrai, Alauddin Javed, Patrick O'Kelly, and Kevin Dick. "Microcell design principles." *IEEE Communications Magazine* 31, no. 4 (1993): 76-82.
- [18]. Qing-An, Z., and A. Dharma. "Handover in wireless mobile networks. Handbook of Wireless Networks and Mobile Computing" Chap. 1, pp. 1–25. 2002. ISBN: 9780471224563, DOI: 10.1002/0471224561, John Wiley & Sons, Inc.
- [19]. Tayyab, Muhammad, Xavier Gelabert, and Riku Jäntti. "A survey on handover management: From LTE to NR." *IEEE Access* 7 (2019): 118907-118930.
- [20]. Chandrashekar, Subramanya, Andreas Maeder, Cinzia Sartori, Thomas Höhne, Benny Vejlgaard, and Devaki Chandramouli. "5G multi-RAT multi-connectivity architecture." In *2016 IEEE International Conference on Communications Workshops (ICC)*, pp. 180-186. IEEE, 2016.

- [21]. 5G New Radio-NR and NG-RAN Overall Description-Stage 2, document 3GPP TS 38.300 Release 15, Version 15.5.0, 2019.
- [22]. Gures, Emre, Ibraheem Shayea, Abdulraqueb Alhammadi, Mustafa Ergen, and Hafizal Mohamad. "A Comprehensive Survey on Mobility Management in 5G Heterogeneous Networks: Architectures, Challenges and Solutions." *IEEE Access*, 2020
- [23]. 3GPP TS 37.340 V15.5.0, "Evolved Universal Terrestrial Radio Access(E-UTRA) and NR; Multi-connectivity," 2019.
- [24]. Manjeshwar, Akshatha Nayak, Pranav Jha, Abhay Karandikar, and Prasanna Chaporkar. "Enhanced UE Slice Mobility for 5G Multi-RAT Networks." In *2019 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pp. 1-6. IEEE, 2019.
- [25]. Rizkallah, Jacky, and Nadine Akkari. "SDN-based vertical handover decision scheme for 5G Networks." In *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, pp. 1-6. IEEE, 2018.
- [26]. Alhammadi, Abdulraqueb, Mardeni Roslee, Mohamad Yusoff Alias, Ibraheem Shayea, and Abdullah Alquhali. "Velocity-Aware Handover Self-Optimization Management for Next Generation Networks." pp. 1354. *Applied Sciences* 10, no. 4, 2020.
- [27]. Farooq, Hasan, Ahmad Asghar, and Ali Imran. "Mobility Prediction Based Proactive Dynamic Network Orchestration for Load Balancing with QoS Constraint (OPERA)." *IEEE Transactions on Vehicular Technology* 69, no. 3, pp 3370-3383. IEEE, 2020.
- [28]. Spoorthi, G. P., and M. B. Akkamahadevi. "Handover Mechanism in 5G mmWave Band." In *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, pp. 772-778. IEEE, 2019.
- [29]. Mouawad, Nadia, Rola Naja, and Samir Tohme. "Inter-slice handover management in a V2X slicing environment using bargaining games." *Wireless Networks* (2020): 1-21.
- [30]. Sun, Li, Jing Hou, and Tao Shu. "Optimal Handover Policy for mmWave Cellular Networks: A Multi-Armed Bandit Approach." In *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6. IEEE, 2019.
- [31]. Lee, Changsung, Hyounjun Cho, Soeun Song, and Jong-Moon Chung. "Prediction-Based Conditional Handover for 5G mm-Wave Networks: A Deep-Learning Approach." *IEEE Vehicular Technology Magazine* 15, no. 1, pp 54-62, 2020.
- [32]. Shayea, Ibraheem, Mustafa Ergen, Marwan Hadri Azmi, Sultan Aldirmaz Çolak, Rosdiadee Nordin, and Yousef Ibrahim Daradkeh. "Key Challenges, Drivers and Solutions for Mobility Management in 5G Networks: A Survey." *IEEE Access* 8 pp. 172534-172552, 2020.
- [33]. Ouali, Kaouthar, Meriem Kassar, Thi Mai Trang Nguyen, Kaouthar Sethom, and Brigitte Kervella. "An efficient D2D handover management scheme for SDN-based 5G networks." In *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1-6. IEEE, 2020.
- [34]. Ko, Haneul, Insun Jang, Jaewook Lee, Sangheon Pack, and Giwon Lee. "SDN-based distributed mobility management for 5G." In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 116-117. IEEE, 2017.
- [35]. Huang, Chung-Ming, Duy-Tuan Dao, and Meng-Shu Chiang. "SDN-FHOR-DMM: a software defined network (SDN)-based fast handover with the optimal routing control method for distributed mobility management (DMM)." *Telecommunication Systems* 72, no. 2 pp. 157-177, 2019.
- [36]. Kaur, Preetjot, and Roopali Garg. "A survey on key enabling technologies towards 5G." In *IOP Conference Series: Materials Science and Engineering*, vol. 1033, no. 1, p. 012011. IOP Publishing, 2021.
- [37]. Jain, Akshay, Elena Lopez-Aguilera, and Ilker Demirkol. "Are mobility management solutions ready for 5G and beyond?" *Computer Communications*, 2020.
- [38]. Akyildiz, Ian F., Jiang Xie, and Shantidev Mohanty. "A survey of mobility management in next-generation all-IP-based wireless systems." *IEEE Wireless communications* 11, no. 4 pp. 16-28, 2004.
- [39]. Kaur, Preetjot, and Roopali Garg. "A survey on key enabling technologies towards 5G." In *IOP Conference Series: Materials Science and Engineering*, vol. 1033, no. 1, p. 012011. IOP Publishing, 2021.
- [40]. Debus, Walter, and L. Axonn. "RF path loss & transmission distance calculations." *Axonn, LLC*, 2006.
- [41]. Alsharif, Mohammed H., Rosdiadee Nordin, Mohammed Mudhafer Shakir, and Athirah Mohd

- Ramly. "Small Cells Integration with the Macro-Cell Under LTE Cellular Networks and Potential Extension for 5G." *Journal of Electrical Engineering & Technology* 14, no. 6: 2455-2465, 2019.
- [42]. Kang, Seongjoon, Siyoung Choi, Goodsol Lee, and Saewoong Bahk. "A dual-connection based handover scheme for ultra-dense millimetre-wave cellular networks." In *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6. IEEE, 2019.
- [43]. Naren, Tada, Patalia Tejas, and Patel Chirag. "Trust appraisal based neighbour defence secure routing to mitigate various attacks in most vulnerable wireless ad hoc network." *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1*. Springer, Cham, 2016.
- [44]. Adedoyin, Mary A., and Olabisi E. Falowo. "Combination of ultra-dense networks and other 5G enabling technologies: A survey." *IEEE Access* 8 (2020): 22893-22932.
- [45]. Xu, Xiaodong, Zhao Sun, Xun Dai, Tommy Svensson, and Xiaofeng Tao. "Modelling and analysing the cross-tier handover in heterogeneous networks." *IEEE Transactions on Wireless Communications* 16, no. 12 (2017): 7859-7869.
- [46]. Lee, Kyoung Seok, and Ae-Soon Park. "Reduction of handover failure for small cells in heterogeneous networks." In *2014 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 707-708. IEEE, 2014.
- [47]. Vasudeva, Karthik, Meryem Simsek, David López-Pérez, and Ismail Güvenç. "Analysis of handover failures in heterogeneous networks with fading." *IEEE Transactions on Vehicular Technology* 66, no. 7 (2016): 6060-6074.
- [48]. "High-level architecture of Mobile Cellular Networks from 2G to 5G." 3G4G, 3g4g.co.uk/Training/intermediate0001.html. [49]. Mousa, Saif Hikmat, et al. "Effective Wide Spectrum Sharing Techniques Relying on CR Technology toward 5G: A Survey." *J. Commun.* 15.2 (2020): 122-147.
- [50]. Chen, Shanzhi, and Jian Zhao. "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication." *IEEE communications magazine* 52.5 (2014): 36-43.
- [51]. Nadembega, Apollinaire, Abdelhakim Hafid, and Tarik Taleb. "Mobility-prediction-aware bandwidth reservation scheme for mobile networks." *IEEE Transactions on Vehicular Technology* 64, no. 6 (2014): 2561-2576.
- [52]. De Rango, Floriano, Peppino Fazio, and Salvatore Marano. "Utility-based predictive services for adaptive wireless networks with mobile hosts." *IEEE Transactions on Vehicular Technology* 58, no. 3 (2008): 1415-1428.
- [53]. Ozturk, Metin, Mandar Gogate, Oluwakayode Onireti, Ahsan Adeel, Amir Hussain, and Muhammad A. Imran. "A novel deep learning driven, low-cost mobility prediction approach for 5G cellular networks: The case of the Control/Data Separation Architecture (CDSA)." *Neurocomputing* 358 (2019): 479-489.
- [54]. Bahra, Nasrin, and Samuel Pierre. "A Hybrid User Mobility Prediction Approach for Handover Management in Mobile Networks." In *Telecom*, vol. 2, no. 2, pp. 199-212. Multidisciplinary Digital Publishing Institute, 2021.
- [55]. Meneghello, Francesca, Davide Cecchinato, and Michele Rossi. "Mobility prediction via sequential learning for 5G mobile networks." In *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1-6. IEEE, 2020.
- [56]. Efroni, Yonathan, Dipendra Misra, Akshay Krishnamurthy, Alekh Agarwal, and John Langford. "Provable RL with Exogenous Distractors via Multistep Inverse Dynamics." *arXiv preprint arXiv:2110.08847* (2021).
- [57]. Sah, Mithila Bihari, Abhay Bindle, and Tarun Gulati. "Issues and Challenges in the Implementation of 5G Technology." *Computer Networks and Inventive Communication Technologies* (2022): 385-398.
- [58]. Series, M. "Guidelines for evaluation of radio interface technologies for IMT-2020." *Report ITU* (2017): 2412-0.
- [59]. Series, M. "Guidelines for evaluation of radio interface technologies for IMT-Advanced." *Report ITU-R M.2135-1* (2009): 1-72.
- [60]. Zhang, Hongtao, and Lingcheng Dai. "Mobility prediction: A survey on state-of-the-art schemes and future applications." *IEEE access* 7 (2018): 802-822.
- [61]. Li, Yun, Bin Cao, and Chonggang Wang. "Handover schemes in heterogeneous LTE networks:

- challenges and opportunities." *IEEE Wireless Communications* 23, no. 2 (2016): 112-117.
- [62]. Mathieu Boutin (2022). Random Waypoint mobility model (<https://www.mathworks.com/matlabcentral/fileexchange/30939-random-waypoint-mobility-model>), MATLAB Central File Exchange. Retrieved July 31, 2022.
- [63]. Kasim, Ahmet Nezih, Ibraheem Shayea, Sajjad Ahmad Khan, Abdulraqueeb Alhammadi, and Mustafa Ergen. "Evolutionary paths towards mobility management in 5G Heterogeneous Networks." In *2020 IEEE Microwave Theory and Techniques in Wireless Communications (MTTW)*, vol. 1, pp. 24-29. IEEE, 2020.
- [64]. Sapkale, Pallavi, and U. D. Kolekar. "Mobility management for 5G mobile networks." *Int J Comput Appl* 182 (2018): 1-4.
- [65]. Akkari, Nadine, and Nikos Dimitriou. "Mobility management solutions for 5G networks: Architecture and services." *Computer Networks* 169 (2020): 107082.