TABLE OF CONTENTS

Page

# LIST OF TABLES

Page

## LIST OF FIGURES

Page

## LIST OF ABREVIATIONS

| | |
|---|---|
| AC | Access Point |
| AC | Admission Control |
| CA | Carrier Aggregation |
| COMP | Coordinate MultiPoint |
| CR | Cell reselection |
| CBDAM | Chance-Based Deferred Acceptance Matching |
| CQA | Channel-Aware Quality of service scheduler |
| CR | Cell reselection |
| CIO | Cell individual offsets |
| CBDAM | Chance-Based Deferred Acceptance Matching |
| c-SON | Centralized-Self-organized network |
| D-eNB | Donor eNB |
| eNB | Evolved Node B |
| FHM | Frequent handover mitigation |
| FRN | Fixed Relay Node |
| FDD | Frequency division multiplexing |
| GPRS | General Packet Radio Service |
| HST | High-speed train |
| HPO | Handover parameter optimization |
| HO | Handover |
| HetsNets | Heterogeneous and small cell Networks |
| IEEE802.11 | Institute of Electrical & Electronics Engineers Local Area Network Standards |
| ISI | Inter-symbol interference |
| IM | Interference management |
| LTE | Long Term Evolution |
| LBEF | Load balancing efficiency factor |
| LSD | Load standard deviation |
| MCS | Modulation coding scheme |

| | |
|---|---|
| MRN | Mobile Relay Node |
| MLB | Mobility load balancing |
| MRO | Mobility robustness optimization |
| MTLB | Multi-traffic load balance |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| OAM | Operation and Management system |
| QoS | Quality of Service |
| QCI | QoS Class Identifier |
| RB | Resource Block |
| RAT | Radio Access Technology |
| RRH | Remote Radio Head |
| RSS | Received Signal Strength |
| RSRP | Reference signal received power |
| RSRQ | Reference signal received quality |
| RRC | Radio Resource Connection |
| SINR | Signal to Interference plus Noise Ratio |
| SNR | Signal-to-Noise Ratio |
| SC-FDMA | Single-Carrier Division Multiple Access |
| TDD | Time division duplexing |
| TTT | Time to trigger |
| UE | User Equipment |
| UMTS | Universal Mobile Telecommunications System |
| UMLB | Utility-based Mobility Load Balancing |
| VPL | Vehicular Penetration Loss |
| VUE | Vehicular User Equipment |
| WI-FI | Wireless Fidelity |
| WIMAX | Worldwide Interoperability for Microwave Access |
| 3GPP | Third Generation Partnership Project |
| 5G | Fifth Generation |

## LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

| | |
|---|---|
| % | Percent Sign |
| dB | Decibel |
| dBm | Decibel -milliwatts (1 mW) |
| GHz | Gigahertz |
| Kbps | Kilo bit per second |
| Km | Kilometre |
| m/sec | Metre per second |
| Mbps | Mega bit per second |
| MHz | Megahertz |
| ms | Millisecond |
| pps | Packet par second |
| Sec | Second |
| Tb | Time per bit |
| $P_{eNB,MRN}$ | The power transmitted by eNB to Mobile Relay Node |
| $G_{eNB,MRN}$ | The gain of the channel between eNB and MRN |
| $P_N$ | Noise power at the relay level |
| $BW_{\ eNB,MRN}$ | Amount of bandwidth assigned to MRN |
| $R_{Tot.}$ | Total throughput achievable |
| $R_{Req.}$ | Required throughput for an application |
| $R_{min.}$ | Minimum acceptable throughput by UEs |
| $PL$ | The path loss between an eNB and an MRN or a user |
| $f$ | Carrier frequency in MHZ |
| $d$ | The distance of MRN or UE from the eNB. |
| $x_{j,i}$ | Association indicator |
| $P_{ji}$ | Preference matrix |
| $RBUR$ | Resource block utilization ratio |
| $N_{s,j}(\tau)$ | Number of resource blocks assigned by small cell $s$ to UE $j$ at period $t$ |
| σ | Standard Deviation of load |

| | |
|---|---|
| $S$ | Number of small cells in a network |
| $M_n$ | RSRP of the neighbouring cell |
| $M_s$ | RSRP of the serving cell |
| $O_{cn}$ | The cell-specific offset of the neighbouring cell |
| $O_{cs}$ | The cell-specific offset of the serving cell |
| $Hyst$ | Hysteresis term for cell $s$ |
| $Off$ | A3 event offset between the serving and neighbouring cells |
| $\varepsilon$ | Set of edge-UEs in a small cell $s$ |
| $TH_{Adp}$ | Adaptive threshold |
| $TH_{INI}$ | Initial threshold |
| $e_s^j$ | Edge-UE |
| $R_{e_s^j}$ | Data rate value demanded by an edge-UE |
| $r_{s,e_s^j}$ | Maximum achievable data rate using the assumed MCS |
| $N_{PRB}^{e_s^j}$ | Number of PRBs required by an edge-UE |
| $\rho(s, e_s^j)$ | Average load to serve the UE $e_s^j$ of small cell $s$ |
| $\hat{\rho}(k, e_s^j)$ | After-handover generated load by edge-UEs |
| $TH_i$ | A4 threshold of serving cell $s$ |
| $\|\|B_s\|\|$ | Neighbouring cells set of serving small cell $s$ |
| $\mathcal{T}_{e_s^j}$ | Set of multiple neighbouring cells of each edge-UE in set $\varepsilon_s$ |
| $\Psi_o$ | Load balancing efficiency factor |
| $X$ | A vector of $n$ criteria |
| $W$ | A vector of $n$ weights |
| $u_j^s(x_i)$ | Utility of a class $j$ UE for a small cell eNB $s$ and a predefined criterion$x_i$ |
| $\mathfrak{U}_{e_o^j}^{\tau^l}$ | UE utility |
| $u_{e_o^j}^{\tau^l}$ | Operator utility |
| $\mathcal{U}_{e_o^j}^{\tau^l}$ | Combined utility |

# INTRODUCTION

Small cells in Long Term Evolution-Advanced (LTE-A) networks include several types such as Pico eNBs, Remote Radio Heads (RRHs), in-band Fixed Relay Nodes (FRNs), Mobile Relay Nodes (MRNs), and Femtocells. Extending the service coverage within macro-cells was the prime objective behind designing small cells; however, they can be densely deployed to increase the capacity of the wireless network significantly (Damnjanovic et al., 2011). Therefore, future networks may adopt the technology of small cells to support ever-increasing data demand. Figure (1) shows a typical scenario of a small cell network. Throughout this research, our work will be focused on both fixed and moving small cell applications.

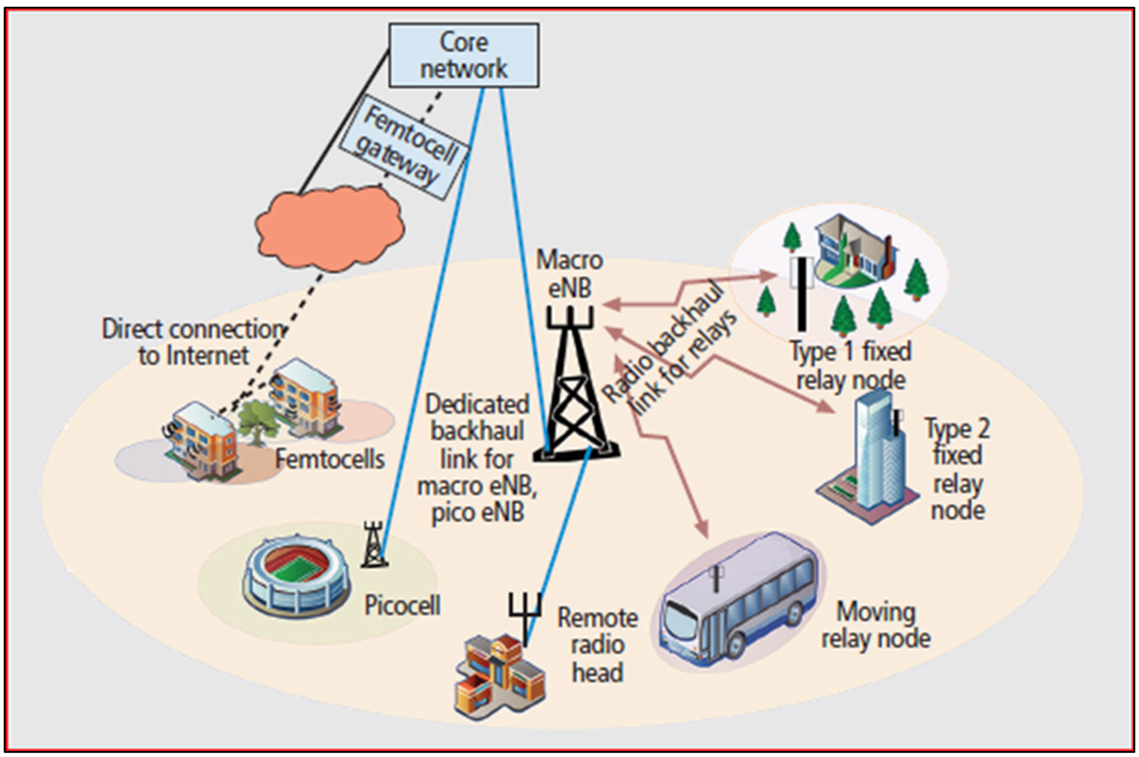



Figure 1 3GPP heterogeneous and small cell networks
Taken from Sui et al. (2013)

On one hand, nowadays, vehicular users, especially transportation passengers, e.g., trains, trams, or buses, represent most wireless internet users. Their number is predicted to increase due to the wide use of smartphones, portable tablets, and laptops. Thus, public transportation

Vehicles will be a data-intensive wireless hotspot. The signals transmitted from the eNB directly to the Vehicular User Equipment (VUE) inside the vehicles are highly attenuated by the Vehicular Penetration Loss (VPL). Thus, more radio frequency power will be transmitted to defeat the VPL and therefore assure an acceptable level of Quality of Service (QoS). Since VUEs will be a great number of broadband subscribers in the near future, it is important to design a new wireless system that guarantees a certain level of QoS with a cost-efficient operation for those VUEs.

In the literature, many solutions have been introduced to serve the VUEs. However, dedicated MRNs deployment might be one of the most effective ways to serve VUEs. MRNs are deployed on top of public transportation vehicles to serve the VUEs, as illustrated in Figure (1). This new technology still needs further study. MRNs are considered as low power nodes that connect to the eNBs via the radio interface. MRNs are surely able to mitigate the Impact of the VPL. Besides, MRN can employ several smart antenna techniques, and advanced signal processing schemes to improve system performance.

Furthermore, MRNs can reuse the FRN physical layer interfaces standardized in the 3rd Generation Partnership Project (3GPP) LTE release 10. An MRN can have its own cell and handle all the layer-2 and layer-3 communication protocols, and thus appears as a regular base station to the VUEs aboard a public transportation vehicle. Multi-RAT functionalities, i.e., LTE, Wi-Fi, GPRS, etc., may be supported by MRN as the access link to VUEs. Moreover, a group of VUEs that are served by the same MRN can be handed over as it is a super-user group. Thereby, the Group Handover helps in reducing the handover failure probabilities. In our thesis, we are going to investigate some techniques that can enhance the VUEs' QoS and performance.

Relay-enhanced LTE is a many-to-one association between Relay Nodes (RNs) and eNB (i.e., several relay nodes can be connected to one eNB but an RN is connected only to one eNB), (Teyeb et al., 2009). Though that is a simple solution to enable relaying in LTE, the end-to-end performance of relayed UEs will be constrained by the capacity available on the backbone link that is accessible through the S1 interface. Radio resources might be enough at the relay link but congestion in the backbone (i.e., on the S1 link) can degrade the performance of relayed UEs. Nevertheless, we may have unfavourable conditions or insufficient radio resources on the relay link (i.e., Un link) with the DeNB despite there are lightly loaded neighbouring cells. Hence, many-to-many connections between RNs and eNBs through the $Un$ interface are introduced to use the sum of the available capacity in all the neighbouring cells. User association is when the user or relay node (in our case) is assigned to different stations in the system. Conventionally, users or relay nodes are assigned to only one eNB at a time. It is proved that the achievable throughput of any terminal depends on the users-eNBs assignment decisions. In this thesis, for a distributed relaying system, a user association method that increases the number of users admitted in the system will be developed (i.e. minimizing blocking). In addition, with the mobility of the relays considered, the handover failure rate will greatly be minimized. Specifically, this work focuses on the first critical problem aforementioned that considerably determines the possibility of admitting a call in the network (Macro cell) without affecting the quality of service required by the call context. A call is rejected if the above condition is not satisfied. The admission control (AC) establishes the rules of acceptance of different call types (new or handover) in the network to ensure a very low block rate and an optimal call acceptance. The challenges in the user association are to minimize the number of blocked VUEs in the network when the network resources are scarce and enhance the QoE of VUEs.

On the other hand, the deployment of residential and non-residential small cells is growing rapidly (Bian et al., 2014). This deployment can be planned or unplanned deployment according to the service operator's policy (Qualcomm, 2019). Unlike a macro network, the low cost of small cells encourages subscribers to install their small cells without any network planning and site-specific system configuration settings. Hence, a significant number of small cells in the network will be randomly distributed.

Mobility of UEs in a small cell network with low service area cells may cause load-imbalance across the cells in the network. The performance of the network in terms of capacity and handover success rate degrades as a result of such an unbalanced load. The shortage of resources in the overloaded small cells leads to poor QoS and increases the handover failure rate when UEs intend to enter those cells though they have lightly loaded neighbouring cells. Consequently, resources of the unloaded cells remain under-utilized though some overloaded neighbouring cells cannot meet the QoS requirements. Thus, the network needs proper configuration and management mechanisms such that the QoS is improved.

System parameters are adjusted manually in the existing networks to reach high levels of operational performance. However, such manual tuning is becoming difficult with the fast evolution of networks. Self-organized network (SON) was introduced to configure, optimize, and heal itself automatically in LTE, and hence decrease the operational complexity (3GPP, 2014.). SON algorithms are categorized into three classes: centralized, distributed, and hybrid. SON has several components such as mobility load balancing (MLB), frequent handover mitigation (FHM), mobility robustness optimization (MRO), and interference management (IM), that help small cells to deliver carrier-grade performance. MLB distributes the UEs load among small cells to enhance the QoS and to increase system capacity. MLB utilizes cells load information to optimize the cell boundaries to offload UEs. SON uses mobility/handover parameters for load balancing (3GPP, 2011; Feng et al., 2008).

MLB distributes the load among the small cells by adjusting the mobility parameters (i.e., handover parameters) according to their load statuses. To shift the candidate UEs, the cell individual offsets (CIO) of the serving and neighbouring cells are adjusted by UEs based on the reported measurements. However, improper handover decisions and offloading sequence for overloaded cells in MLB might cause an inefficient usage of resources or degrade the service.
The performance metrics that define the QoS required by an application or the network are as following:

- Throughput: the number of packets received per time unit represents the throughput which might be normalized by dividing the receiving packet rate over the sending packet rate.

The objective is to improve the throughput for UEs served by an MRN or a fixed small cell.

- Packet Loss rate: this can be defined as the percentage of the packets that are lost in the network due to congestion, or link failure. The objective is to minimize the Packet Loss rate.
- Delay: two types of delay are defined in any network: the end-to-end delay and the handover delay. The key objective is to minimize the delay as much as possible.
- Blocking probability: it is defined as the probability that some UEs cannot achieve the minimum required data rate in the network. The objective is to minimize the blocking probability as much as possible.
- Standard deviation: The range of Standard deviation is in the interval [0, 1], with a lower value representing a highly balanced load distribution amongst all active small cells. Therefore, minimizing is one of the objectives to achieve a highly balanced load in small cell networks.
- The number of handovers: Mobile UEs may experience regular or forced handovers among small cells based on their speed and session durations and the load balancing method adopted. Frequent or redundant handovers cause packet losses and delays and consequently leads to a poor UE's QoE. The objective is to minimize the number of handovers as much as possible when realizing a balanced-load network.

Optimization of all or some of the factors depends on the type of application (i.e. its required QoS). For instance, the delay must be optimized for delay-sensitive traffics as the case for voice/video applications while some packet loss is tolerated. However, the data traffic can tolerate delay, but the packet loss rate is sub-optimized.

**1. Motivation**

The increases in the use of smartphones and applications for information and communications technologies are causing a rapid increase in the demand for mobile broadband services with higher data rates and higher quality of service (QoS). According to the Cisco Visual

Networking, the expected global mobile data traffic in 2022 is 77 exabytes, which is a seven-fold increase over 2017 (Cisco, 2019). As a result, mobile networks need serious steps to accommodate this massive traffic growth. The small cell is a key part of the fifth-generation (5G) network to support the forecasted data demand and enhance the network capacity (Hoadley et al., 2012). This technology has many advantages that can be summarized as follows:

- A small cell is a low power, cost-effective radio-access point with low service areas ranging from tens to several hundred meters (Forum, 2014).
- Small cells can improve indoor/outdoor coverage and network capacity by reducing the path loss and reusing the whole spectrum.
- Low operational cost due to its self-organization feature that configures, optimizes, and healss itself automatically without human intervention.
- A small cell network gives a friendly network with a low transmission power environment.
- Saving the energy of the base station and the life of batteries.
- Better performance can be provided by reducing the load of the cellular network.

As a result, providing QoS in this technology is a very significant topic that needs to be addressed. It is proved that the achievable throughput of any terminal depends on the users-eNBs assignment decisions. That motivates us to study the mobile user association to stationary and moving relays to assure higher throughput, and less blocking rate and packet loss.

The mobility of users across low service areas of the small cells can lead to an unbalanced load status across the network. Furthermore, an overload state may occur during the small cell selection procedure. Hence, the network performance degrades in terms of call admission rates, handover success rates, and capacity. When users are moving into a cell request data rates that exceed the cell capacity, even though the cell has underloaded neighbouring cells, the cell becomes congested and this causes either handover failures or poor QoS. Accordingly, overloaded cells cannot deliver the QoS requirements, whereas other neighbour cells' resources remain underutilized. Therefore, this motivates us to find a better configuration for the network and management to avoid this issue.

## 2. Problem Statement

The most important issues in this context are the Quality-of-Service (QoS) support, resource and mobility management in 5G Small Cell Systems. Due to the nature of small cells, many potential problems will arise, thereby the QoS of UEs will be degraded. Some issues have been widely studied in the literature. However, many aspects have not been considered in those studies. The issues that need to be addressed for fixed and mobile small cell networks are the following:

- **Service unavailability:** caused by the rejection of calls due to lack of resources or lack of network coverage (dead zones). Service providers have tried to correct the problem by expanding their network coverage and increasing the ability of cells, but the problem still posed.
- **A decrease in throughput:**
  - UEs moving into an overloaded cell, whose resources are depleted, will either be dropped or suffer serious low throughput.
  - Handing over a large group of vehicular UEs into a target eNB or a fixed small cell will disturb their resource allocation. Furthermore, the problem becomes worse when several MRNs move through a network that is overlaid with many fixed small cells.
- **Service interruption:** The frequency of handovers and delay can rapidly increase with the deployment of small cells as the mobile UEs move across them. As a result, that can rapidly generate a high packet error rate (Packet Loss) and a high blocking probability rate especially for real-time traffic (such as streaming services).
- **Imbalance and network overload:** Mobility of UEs in a small cell network with low service area cells may cause load-imbalance across the cells in the network. The performance of the network in terms of capacity and handover success rate degrades as a result of such an unbalanced load. The shortage of resources in the overloaded small cells

leads to poor QoS and increases the handover failure rate when UEs intend to enter those cells though they have lightly loaded neighbouring cells. Consequently, resources of the unloaded cells remain unutilized though some overloaded neighbouring cells cannot meet the QoS requirements. However, performing load balancing for an overloaded network incurs other costs such as an increased number of handovers and a higher link failure rate if cell individual offsets are not set properly.

## 3. Research Objectives

The main objective of this thesis is to develop a new user association and mobility load-balancing scheme for small cell networks. The user association scheme enables the operator to maximize the number of admitted UEs into the system by utilizing multiple backhaul links for mobile relays.

The MLB algorithm improves the QoS in the network by considering UE and operator utilities. Furthermore, MLB algorithm operation includes a new technique that specifies the sequence of overloaded cells during algorithm operation to maximize QoS. Moreover, The MLB minimizes the number of handovers during the offloading process. The resulting work should solve the following main issues:

- How to maximize the number of UEs admitted into the system;
- How to maximize total user satisfaction in the system;
- How to maximize total operator satisfaction in the system;
- How to improve handover performance in the system; and
- How to improve load balancing efficiency in the system.

Different models will be developed to solve these issues. Then, the output of the proposed models will be compared to those of others found in the literature.

## 4. Methodology Overview

In our methodology, we designed a dynamic matching model-based user association algorithm that considers the real-time variation of the wireless channel of MRNs, as shown in Figure 2. A modified deferred acceptance algorithm is used to perform the matching negotiation between eNBs and relays, as well as UEs. It is a matching between two sets (eNBs and UEs) with the elements of UEs' set proposing and the elements of eNBs set responding.

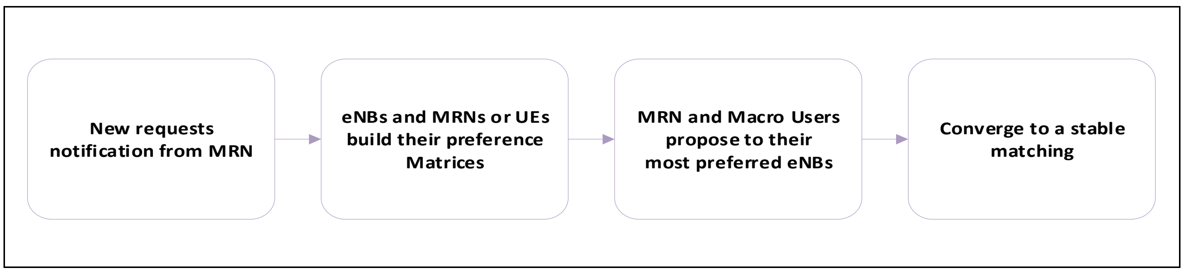



Figure 2 Our Proposed Matching Game Algorithm in LTE-A network

On the other hand, a very important aspect should be considered to improve the QoS, which is balancing the load in our network after the user association phase. The shortage of resources in the overloaded small cells leads to poor QoS and increases the handover failure rate when mobile UEs intend to enter those cells. Our proposed algorithm will distribute the UEs load among small cells based on user and operator utilities to enhance the QoS and to increase system capacity.

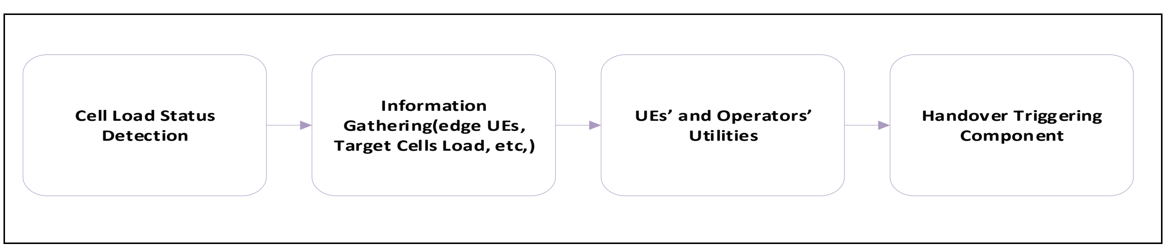



Figure 3 Our Proposed Mobility and Resource Management

Figure 3 shows the offloading process that takes place by handing over each candidate of edge-UEs of the overloaded cell to the best neighbouring cell by calculating the aggregated utility for each neighbouring cell. The aggregated utility is a function in the operator utility and the user utility. The operator utility is calculated for each potential handover based on the load of the neighbouring small cells. On the other hand, the user utility calculation is based on the sigmoid function by considering different criteria (e.g., delay, data rate, etc.) for each edge-UE involving a handover process.

However, the cost of our proposed MLB is a higher number of handovers. Thus, we proposed another method that balances the network load with a minimum number of handovers.

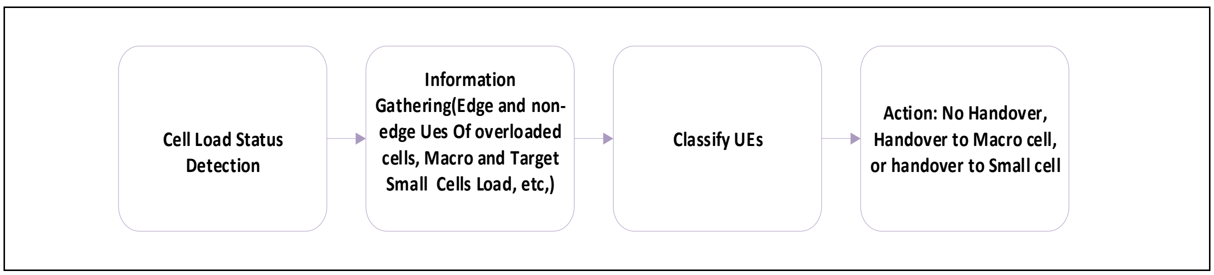



Figure 4 Our Proposed Enhanced Mobility and Resource Management

In this algorithm, which is depicted in Figure 4, we classify UEs of the overloaded cells as very-slow edge, slow edge or fast edge and non-edge UEs. Then, the algorithm excludes the very-slow edge UEs. However, it transfers the slow edge UEs to the neighbouring small cell and the fast edge and non-edge UEs to the Macro cell. As a result, unnecessary handovers are avoided, and the frequent handover is minimized.

## 5. Thesis Contribution

In this thesis, three main contributions are introduced. The first contribution is the designed matching game-based user association scheme for solving the assignment problem for new or handed-over UEs, the algorithms proposed in the literature consider association to base stations with the minimum path-loss values that makes the process not efficient and the UE could notice a severe degradation of QoS. Our algorithm increases the number of admitted users in the MRN system under resource constraints and QoS for both new and handed-over calls. This performance is a result of resource availability at eNBs, algorithm capability to shift users to neighbouring lightly loaded cells and user (or MRN) locations.

The second contribution is the proposition of the Utility-based Mobility Load Balancing algorithm (UMLB) by considering both the operator utility and the user utility at the same time for each lightly-loaded neighbouring cell. The main reason for choosing the utility function is that it expresses the satisfaction of any metric with a numerical value (e.g., cell load, reference signal received power (RSRP), or reference signal received quality (RSRQ)). Therefore, it is easier and less complex to compare these numerical values together to obtain the best decision during a handover process compared to the heuristic algorithms used in (Fedrizzi et al., 2016). The algorithm starts by determining the edge-UEs of an overloaded cell that needs to be offloading to lightly loaded neighbouring cells. The offloading process takes place by handing over each candidate of edge-UEs of the overloaded cell to the best neighbouring cell by calculating the aggregated utility for each neighbouring cell. The aggregated utility is a function in the operator utility and the user utility. The operator utility is calculated for each potential handover based on the load of the neighbouring small cells. On the other hand, the user utility calculation is based on the sigmoid function by considering different criteria (e.g., delay, data rate, etc.) for each edge-UE involving a handover process. Furthermore, we introduced a new term named load balancing efficiency factor (LBEF) that considers a load of neighbouring cells and the edge-UEs for each overloaded cell. This factor specifies the sequence of overloaded cells for the UMLB algorithm operation.

The third contribution is the Utility-based Mobility Load Balancing algorithm with Handover minimization (UMLB-HO) by considering not only the edge UEs but also the non-edge UEs for overloaded small cells during the load balancing process. The UMLB-HO algorithm is just an extension of the UMLB algorithm but with the aim to balance the load among the network with the minimum number of required handovers. The main objective for any overloaded small cell during the MLB is to determine the fast-moving non-edge UE and transfer them to the underloaded Macro cell. Moreover, the small cell will determine the fast-, slow or very slow-moving edge-UEs during the MLB and transfer them to either under loaded neighbour small cell or Macro cell. We defined four important terms to specify whether the UE is fast-, slow or very slow-moving in order to determine the best handover decisions for each UE. If the UE is a non-edge fast-moving, UE is going to be handed over to the Macro cell instead of small cells to avoid the unnecessary frequent handovers among the small cells. However, if the UE is an edge fast-moving, the UE will be handed over to the macro cell to avoid unnecessary frequent handovers. Moreover, if the UE is an edge slow-moving, he will be handed over to a neighbour small cell. While if the UE is a very slow-moving edge UE, he will not be handed over because his session will end before he leaves the serving cell, thus unnecessary handover is avoided.

## 6. Thesis Outline

The rest of the thesis is organized as follows: in Chapter 1, we present some background information about resource and mobility management, and in more details the mobility load-balancing process from the literature. In the same chapter, related work is discussed about mobility management and load balancing techniques proposed by other researchers.

In chapter 2, the performance of the dual-backhaul mobile relay system was evaluated in terms of average user throughput, packet loss and outage probability. In Chapter 3, our proposed admission control for a dual-backhaul 5G MRN system is presented along with problem formulation, and simulation results. In chapter 4, a dynamic mobility load-balancing algorithm is introduced, which is based on utility functions. Algorithms, results, analysis and discussions are also introduced. Chapter 5 introduces an enhanced dynamic mobility load-balancing algorithm that targets the increased number of handovers. Problem formulation, algorithm, simulation and results are introduced. Finally, we conclude our work and discuss possible research directions in Chapter 6.

# CHAPTER 1

# RELATED WORK AND BACKGROUND INFORMATION

## 1.1 Introduction

In this chapter, we explain some background information about the different resource and mobility management techniques available and we introduce the literature review on user association and mobility load balancing in small cell networks.

## 1.2 Related Work on Resource and Mobility Management

In this section, we will present some existing techniques in the literature for resource and mobility management for 5G small cell networks.

### 1.2.1 Mobility Management

The main objective of mobility management in 5G small cell Networks is to assure the continuity of the service during the handover process. Handover is the most sensitive point in the convergence of any two adjacent cells. This transition is assumed homogeneous and transparent to UEs, which implies that the mobile UE must be auto-configured with the new settings without user intervention. Figure 1.1 shows the types of handover: horizontal and vertical handovers.

**Horizontal Handover**

When the network changes, but not the technology, such as a transfer from cell 1 to cell 2.

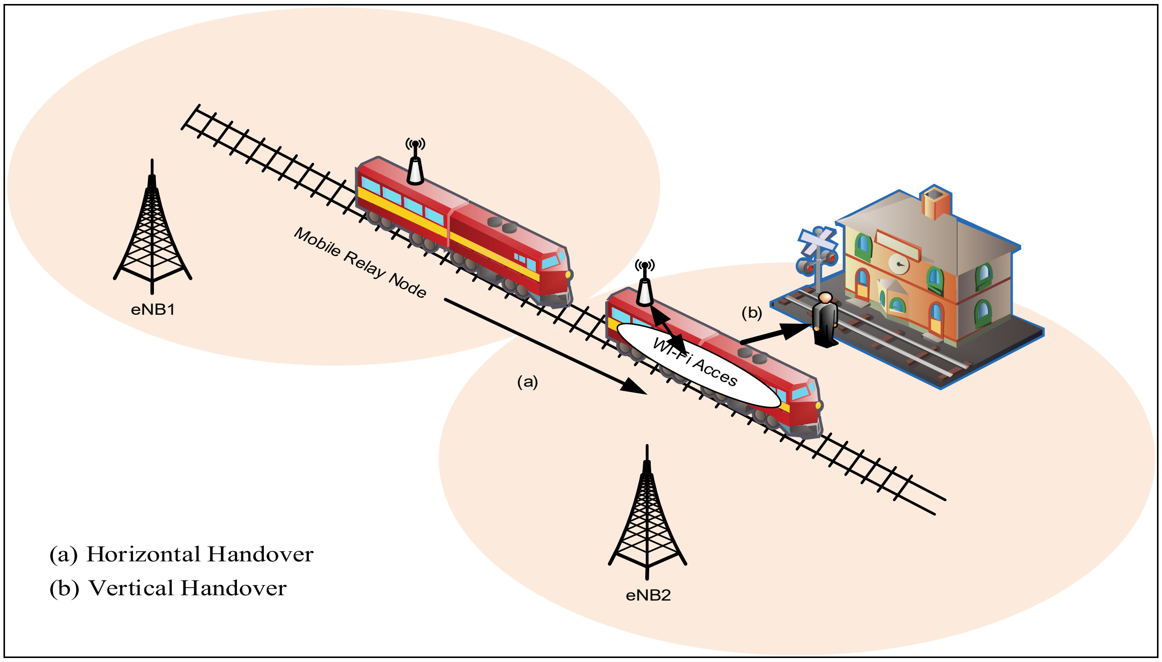



Figure 1.1 Horizontal and Vertical Handovers

One of the challenges that limits the performance and reliability of small cells system is the handover during high-speed movement. In other words, passing multiple cells within several seconds requests a massive number of handovers. Before initiating those handover procedures, the UE has to perform cell measurement and report, Radio Resource Connection (RRC) reconfiguration random access as soon as it enters a crossed cell. Thus, it was proposed that the users boarding a high-speed vehicle are handed over as a whole group. Therefore, system performance is enhanced in terms of mobility and resource management (Li et al., 2012; Luo et al., 2012).

A novel technique to avoid handover problems in LTE has been proposed by authors in (Zhang et al., 2014), which is based on collaboration among macro-femtocells grouped according to nearby base stations. Each group pre-fetches higher layer packets to reduce the latency in handover process.

Another study (Zhao et al., 2011) introduced a new handover algorithm for mobile relay stations to improve the handover success rate. The algorithm is based on the relative speeds of the UEs to the serving DeNB and the target DeNB.

Many Handover schemes have been proposed to solve the problem associated with the frequent handover of mobile equipment in a high mobility circumstance. One of the optimized handover processes for LTE network is designed based on the coordinated multiple point (ComP) transmission technology and dual-vehicle station coordination mechanisms (Luo et al., 2012). The underlying technique of this optimized scheme makes use of the diversity gain in the overlapping area of two adjacent eNBs as the UE moves from one to another. Because of the high level of diversity gain, the probability of handover failure decreases and the Quality of Service (QoS) improves because of the high level of reliability (Pan et al., 2014). Such a technique can be useful for fast-moving UEs because it reduces the probability of service interruption. Then, this technology has been extended to Mobile Relay systems to ease the group handover and decrease the service disruption time. Moreover, the large number of UEs on-board of a bus or train may require a large number of resources that cannot be provided by one backhaul link. Still, this issue has not been investigated in the literature from the resources point of view.

**Vertical Handover**

In the case when the UE moves from a network to another one with different technology, for example, a transfer from LTE cell to Wi-Fi access–supported Relay Cell and vice versa. Several pieces of research have been done in this area in order to achieve an optimal vertical handoff algorithm that allows an inter-cell transfer between any two heterogeneous networks, e.g. LTE and Wi-Fi, in minimal delay and without service interruption. Not only the IP address is changed by switching the network connection, but the network interface also changes. Mechanisms such as mobile Ipv6 and Ipv4 have been proposed to solve the IP address changing problem (Nasser et al., 2006).

Executing handover may consider two main techniques that are inherent to handover (Cho et al., 2005):

- **Hard Handover:** the mobile UE first disconnects from the first cell and then reconnects to the target DeNB.
- **Soft Handover:** the mobile UE connects to the target DeNB first and then disconnects from the old DeNB.

The main parameters used in the literature for the vertical HO decisions and solutions are:

- Availability (RSS, SNR, SINR ...)
- The available bandwidth.
- Latency.
- Reliability.
- Energy consumption.
- Application Type (real-time or non-real time).
- Security.

In fact, several ways of research have been developed only for the static scenario (no mobility) with a focus on maximizing throughput across both networks while minimizing the number of handovers and eliminating the ping-pong effect.

We present below some vertical handover algorithms which characterized by the decision based on different metrics such as user preference, cost, resources of both networks, signal strength, and finally moving speed of the mobile.

Many Vertical handover (HO) algorithms have been proposed in the literature. In (Van Quang et al., 2010), authors proposed an algorithm for HO decision using the metric of Received Signal Strength (RSS); however, using RSS in heterogeneous networks does not give good results. Furthermore, articles (Bing et al., 2003; Lv et al., 2008) combined other metrics with RSS, such as distance between UE and eNB antennas, and the service cost. But the algorithm becomes more complex as well as excessive delay and high power (Yang et al., 2007). Similarly, in (Zhu et al., 2006) cost and/or speed of movement of mobile users have been used as the main indicators and RSS algorithm as a secondary metric. This approach brings better results in terms of rates, cost and blocking probability.

Work in (Chou et al., 2006) used the signal-to-noise ratio (SNR) and traffic type as the metrics for the HO decision. Its goal was to maximize the throughput of the network and minimize the ping-pong effect. Lin, H et al. proposed a QoS-Based Vertical HO. In addition, reference (Yang et al., 2007) uses the combined effects mentioned above including signal-to-interference-noise Ratio (SINR) to make HO decisions for multi-attribute QoS considerations. Still, all the above-mentioned proposed techniques were studied from the core network point of view; however, integrating Wi-Fi in RAN makes it a different issue that needs to be investigated in terms of mobility and resource management.

### 1.2.2 Resource management

Channels in terms of frequency, time intervals (time slots), transmission power and battery power represent the resources in the wireless systems. Good resource management can help service providers reduce costs, increase revenues while ensuring a better QoS. Resource management can improve outcomes in terms of handover blocking probability and ensure continuity of service. The main issue which mainly affects the handover is admission control, bandwidth reservation, and the scheduling policy (Van Quang et al., 2010). Figure 1.2 shows these important components.

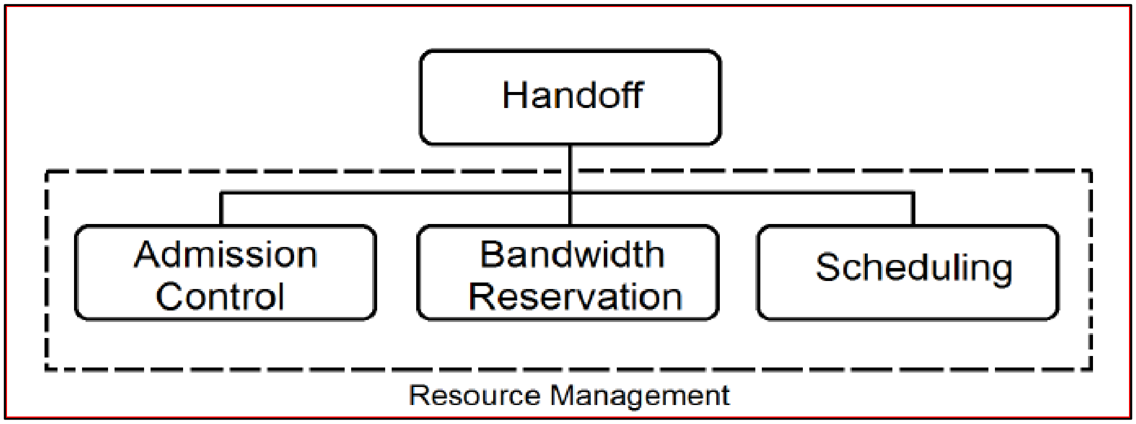



Figure 1.2 Resource management in connection with the handover

#### 1.2.2.1 Admission Control

The admission control (AC) establishes the rules of acceptance of different call types (new or handover) in the network to ensure a very low block rate and an optimal call acceptance. Therefore, the purpose of admission control is to determine the possibility of admitting a call in the network (Macro cell) without impacting the QoS required by the call context. A call is rejected if the above condition is not satisfied.

To do this, Admission Control establishes a priority between new calls (lowest priority) and active calls (highest priority). From the UE experience perspective, the call interruption rate is more critical because it is less pleasant to lose a call that is in progress. Various policies have been proposed in the literature for the admission control in homogeneous networks that can also be applied in the case of a heterogeneous network. In these researches, the AC mechanisms have been based on the quality of the received signal, the available bandwidth, the impact on ongoing QoS, priority of UEs, revenue ... etc. Research done in (Yu et al., 2002) has presented an admission control policy based on mobility. The idea is to rely on the mobility-related information to estimate future resource needs that the UE will need in each of the neighbouring cells. Similar policies have demonstrated better results by combining mobility with the estimation of resource usage according to the duration of the call (Nicopolitidis et al., 2003).

Other policies based on revenue have been proposed to maximize the new call acceptance which means maximizing the revenue since each new call is a potential source of income. In (Nelakuditi et al., 1999), a resource allocation strategy was presented to minimize the call blocking rate when receiving a new call in a saturated cell (with no available resources). The strategy is to move a current call to a neighbouring cell in order to release resources currently used and reallocate it to the new demand. In LTE standard (Access, 2009), the user association scheme is based only on the Received Signal Strength Metric (RSS). However, that scheme was not designed for a heterogeneous network. In (NTT, 2010) a bias is added to the Reference Signal Received Power (RSRP) to extend the small cell coverage. In (Qualcomm, 2010), the user will choose the station that guarantees the minimum path loss values. In (Saad et al., 2014), based on matching game theory, the uplink user association scheme is proposed to solve the assignment problem for small cell networks. However, this scheme considers the fixed small cell networks and only the access link between users and various eNBs. The admission control strategy in the Mobile Relay is different from those have been used for regular small cells since it acts as a super-user group.

#### 1.2.2.2 Bandwidth Reservation

Bandwidth in a wireless network is a valuable and important resource. The handover is successfully executed if the bandwidth is available/reserved in the target network/cell. For this, the simplest solution is to reserve a fraction of the bandwidth in each network (Forum) for the handover calls only. However, the difficulty is to find how much bandwidth should be reserved to minimize the handover block rate while maximizing the use of overall network bandwidth. Much research has been proposed to dynamically manage the resource allocation in terms of bandwidth such as: sharing the total bandwidth by all traffic classes or dividing the bandwidth into separate parts for each traffic class (Diederich et al., 2005).

#### 1.2.2.3 Scheduling

The scheduling algorithm provides a resource-multiplexing mechanism among UEs at every time instant. It allows the packet switching mode support on the radio interface. Sharing these radio resources is mainly based on the channel state, the packet delay and throughput required for each UE. However, an optimal scheduling strategy must share resources among UEs in a way that provides an equitable level of QoS while optimizing resource usage.

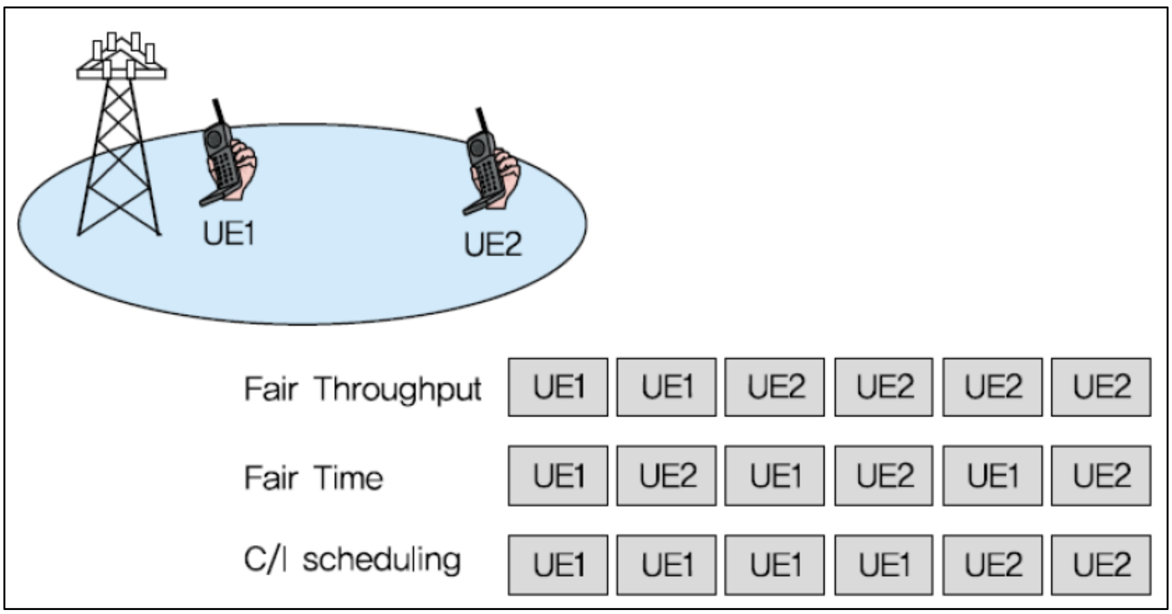



Figure 1.3 The Three Conventional Scheduling Algorithms
Taken from NSN: LTE MAC/RLC/PDCP/RRC (2011)

The best-known scheduling algorithms in the literature depicted in Figure 1.3 are as follows (Dahlman et al., 2013):

- Round Robin Algorithm (also known as Fair Time scheduler): also called “Fair Time scheduler” that share resources equitably among UEs without considering the radio channel state.

- C/I scheduler: this one considers the radio channel state and primarily seeks to maximize the radio resource efficiency without considering fairness among UEs.
- The Fair Throughput algorithm: it provides a fair rate for all UEs even if the resource usage is far from optimal.

The introduction of the Mobile Relay technique certainly may involve new scheduling algorithms. Also, the scheduling strategy is closely related to other functions of resource management such as admission control and bandwidth reservation.

### 1.2.3 Mobility Load Balancing

The mobility parameters can be self-optimized to the cell current load and in the neighbouring cells to enhance the capacity of the system. Also, human intervention is significantly minimized. The QoS experienced by UEs with load balancing should not be worse than that with regular mobility. The main objective is to deal with the uneven traffic load by optimizing the reselection/handover parameters. Nonetheless, the number of handovers required to do the load balancing should be kept as minimum as possible.

Part of UEs at the edge of the congested cells can reselect or handover to the less loaded cell. Hence, the load of the cell is balanced, and the system capacity is enhanced. Figure 1.4 shows the actions needed to perform mobility load-balancing optimization.

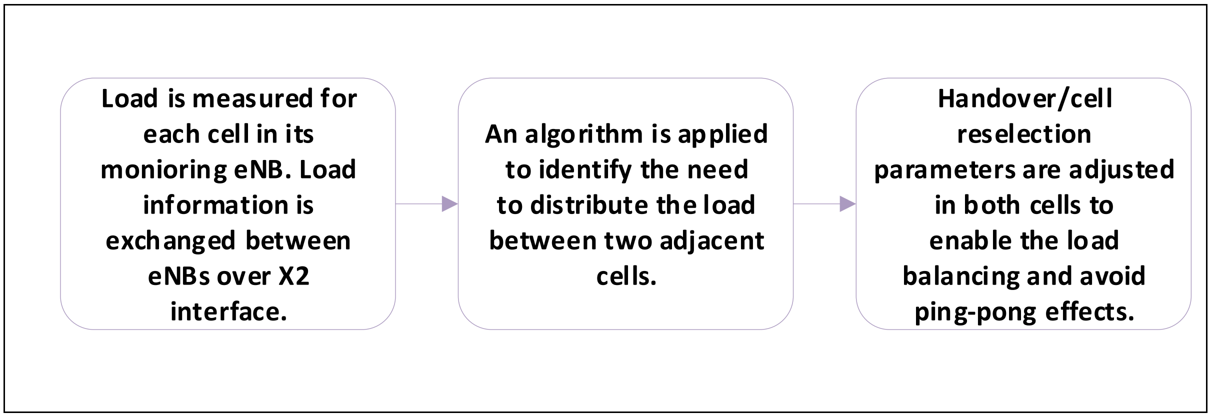



Figure 1.4 Mobility Load Balancing Procedure

The definition of the load has many aspects. Some models consider the radio load, transport network load or the processing load. The radio load might be split between uplink load and downlink load or split between different QoS Class Identifiers (QCIs). Based on the defined load, the algorithm distributes the load across the network. An algorithm is found to determine when and how the load is balanced. In other words, how the overload status is detected and handled.

Mobility load balancing modifies the handover parameters to control the overload situation. The two main parameters that might be used are the Hysteresis and Cell Individual Offset (CIO). For the sake of load balancing cells should use the CIOs if they want to steer traffic to certain neighbours and not all of them. When a cell modifies the CIO, only one neighbour cell would have to adjust its corresponding CIO. Adjusting hysteresis requires to require a modified hysteresis in all neighbour cells thereby causing these changes to ripple through a large part of the network. Moreover, to avoid ping pongs, it is necessary to adjust the mobility parameters in both the source and target cells.

Information exchanged among cells over the X2 interface to make the two-sided change when balancing the load. The change of CIO must be within a specific range defined by the cell coverage overlap of the two cells. If the change is made out of this range, HO failures and call drops occur.

Intra-LTE mobility parameters auto-adjustment based on the current load of the small cell network can enhance the system capacity compared to the static/non-optimized cell mobility parameters (Feng & Seidel, 2008). However, the UE QoS shall not be affected negatively with the forced load balancing.

Researchers in (Kwan et al., 2010) was the first to demonstrate through simulation the effectiveness of simple load balancing algorithms in reducing the call blocking rate and increasing cell-edge throughput based on auto-adjustment of handover parameters.
In (Lobinger et al., 2010), based on RSRP measurements and load of neighbouring cells, overloaded cells group UEs according to the best neighbouring eNB, and for each handover offset value, it sorts the neighbouring eNBs in descending order concerning the number of possible handovers. A whole group will be handed over if their predicted load does not exceed the acceptable level at neighbouring eNB.

MLB algorithm proposed in (Zia et al., 2013) considered non-adjacent neighbourhood cells in the optimization area. The radio link condition of neighbouring cells is taken into consideration when offloading UEs from the source cells.

MLB optimization and handover parameter optimization (HPO) algorithms influence the handover decisions of the UEs. That interaction reduces the desired effects of each function. The coordination between MLB and HPO is investigated in (Lobinger et al., 2011). The coordinator provides a solution that combines the strengths of the individual algorithms to better performance.

In (Huang et al., 2015), a multi-traffic load balance (MTLB) algorithm is presented to balance the traffic load and improve the network capacity with an appropriate handover procedure. A new cell selection is adapted to enhance the quality of service of UEs. Besides, the handover threshold and TTT (time to triggering) are adaptively adjusted to reduce the call drop rate with a more balanced-load network. Two conditions are accounted for the handover procedure: The signal strength condition and the RB condition. That helps in avoiding the wrong eNB or unnecessary handovers.

In (Rajpoot et al., 2018), researchers presented an MLB algorithm and discussed its properties in terms of costs and gains. Their objective is to minimize the load standard deviation (LSD) to distribute the load evenly over the network. Simulation results show that the MLB algorithm can reduce the LSD significantly. However, the algorithm required more handovers than the no-MLB operation, and the number of RLF produced was higher as well.

Researches in (Oh et al., 2016), proposed an MLB algorithm based on cell reselection (Vu et al.) which works in accordance with the Mobility Robustness Optimization (MRO) function. In other words, when UEs are in the radio resource control (RRC) idle mode, the algorithm adjusts the CR parameters to make them camp on the lightly loaded cell. Once UEs switch to the (RRC) connected mode, they will belong to the lightly loaded cell selected in the idle mode.

Not only mobility of UEs can be utilized in load balancing, but also there are different types such as coverage and capacity optimization (Yamamoto et al., 2012). When a small cell is detected to be overloaded, the SON has a function that decreases the power and hence makes some edge-UEs offload to the lightly loaded side of the network.

The mobility load balancing is performed by adjusting the mobility parameters; thus, the impacts of user mobility on the algorithm have been studied. Those impacts are evaluated and compared through computer simulations (Oh et al., 2018). Results show that MLB algorithms realize higher gain for pedestrian users with circle mobility and vehicle users in rectangle mobility.

In general, the previous work addressed the MLB problem by mainly considering the operator preference without considering the UEs preferences during the handover process. Moreover, the introduced algorithms in the literature did not follow a proper sequence in ordering the overloaded cells in the MLB problem. On the other side, previous research has never considered the increased number of handovers caused by MLB algorithms. An excessive number of handovers that may degrade the overall network performance. This thesis introduces algorithms that integrate the UE utility and operator utility during load balancing in the network. Furthermore, it presents an LBEF term that is used to specify the sequence of overloaded cells for the MLB algorithm operation. Finally, a new algorithm is proposed to mitigate the handover cost of the MLB algorithms.

# CHAPTER 2

## DUAL-BACKHAUL LINKS IN LTE-A MOBILE RELAY SYSTEM FOR HIGH-SPEED RAILWAYS

### 2.1 Introduction

Initially, the relaying functionality was supported to extend the LTE radio access technology where the user equipment (UE) communicates via a relay node which is wirelessly linked to the eNB (called a Donor eNB) (Parkvall et al., 2011). The desire of high-Speed public transportation passengers to have access to the internet with high-speed data services leads to adopting LTE-A due to its good performance (Atat et al., 2012). Hence, in release 12, Mobile Relay Nodes (MRN) has been proposed to extend the cellular coverage. In addition, MRN can effectively avoid Vehicular Penetration Loss (VPL) (3GPP, 2012b; Alsharoa et al., 2014). With the presence of moderate to high VPL, the performance of mobile relays has been shown to outperform the performance of the fixed relay as well as the direct links between eNB and UEs. The work in (3GPP, 2012b) has shown that the performance of high-speed vehicular systems with MRN is better than that without Mobile Relay Nodes assistance.

One of the challenges that limits the performance and reliability of LTE system is the handover during high-speed movement (Li et al., 2012; Luo et al., 2012). A novel technique to avoid the handover problem in LTE has been proposed by authors in (Zhang & Wang, 2014), which is a collaboration among macro-femtocells grouped according to nearby base stations. Each group pre-fetches higher layer packets to reduce the latency in the handover process. Many Handover schemes have been proposed to solve the problem associated with the frequent handover of mobile equipment in high mobility circumstances. One of the optimized handover processes for LTE network is designed based on the coordinated multiple point transmission technologies (CoMP) and dual vehicle station coordination mechanism (Luo et al., 2012). The underlying technique of this optimized scheme makes use of the diversity gain in the overlapping area of two adjacent eNBs as the UE moves from one to another. Because of the high level of diversity gain and reliability, the probability of handover failure decreases and the Quality of Service

(QoS) increases (Pan et al., 2014). Such a technique can be useful for fast-moving UEs because it reduces the probability of service interruption. Then, this technology has been extended to Mobile Relay systems to ease the group handover and decrease the service disruption time. Moreover, the large number of UEs on-board a bus or train may require a large number of resources that cannot be provided by one backhaul link. Still, considering multiple backhaul links schemes to provide the Mobile Relay Node with more bandwidth has not been fully investigated. In this research, in order to maximize the UE average throughput, facilitate the group handover, and improve the UEs' Quality-of-Service, we propose a new architecture that adopts dual-backhaul links (A Carrier Aggregation Design). As the UEs are served by different providers, the dual-backhaul links would be from divers-operator eNBs. So, cooperation among the operators is assumed within our model.

## 2.2 System model

The system model considered is a deployment of intensely overlapped macro eNBs that thoroughly cover the geographical region, with a Mobile Relay Node (MRN) installed on top of the roof of a high-speed train (HST) that carries up to 60 UEs. The communications between VUEs and eNBs are performed over two hops via the MRN. MRN is normally served by various providers, operators A and B, as shown in Figure 2.1. The UEs' data is assumed to be real-time traffic.

Since the eNBs belong to different operators, they will certainly overlap with each other. We assume that the trains are operated at high speed on a known straight trajectory. Essentially, the Mobile Relay Node (MRN) – eNBs transmission operations are transparent to the UEs hosted on the train. In this work, the MRN serves the users via two backhaul links from two distinct eNBs, namely primary eNB (P-eNB) and secondary eNB (S-eNB). The primary and secondary links are selected based on the received signal strength (RSS) at the MRN, i.e. the secondary Link signal power strength is lower than that of the primary link. Accordingly, at any time, a handover of the Secondary Link to a target eNB is to be performed. The benefits of the dual-backhaul links for the mobile relay node are presented in terms of improving the user Throughput, decreasing the Packet Loss, and minimizing the Blocking Probability.

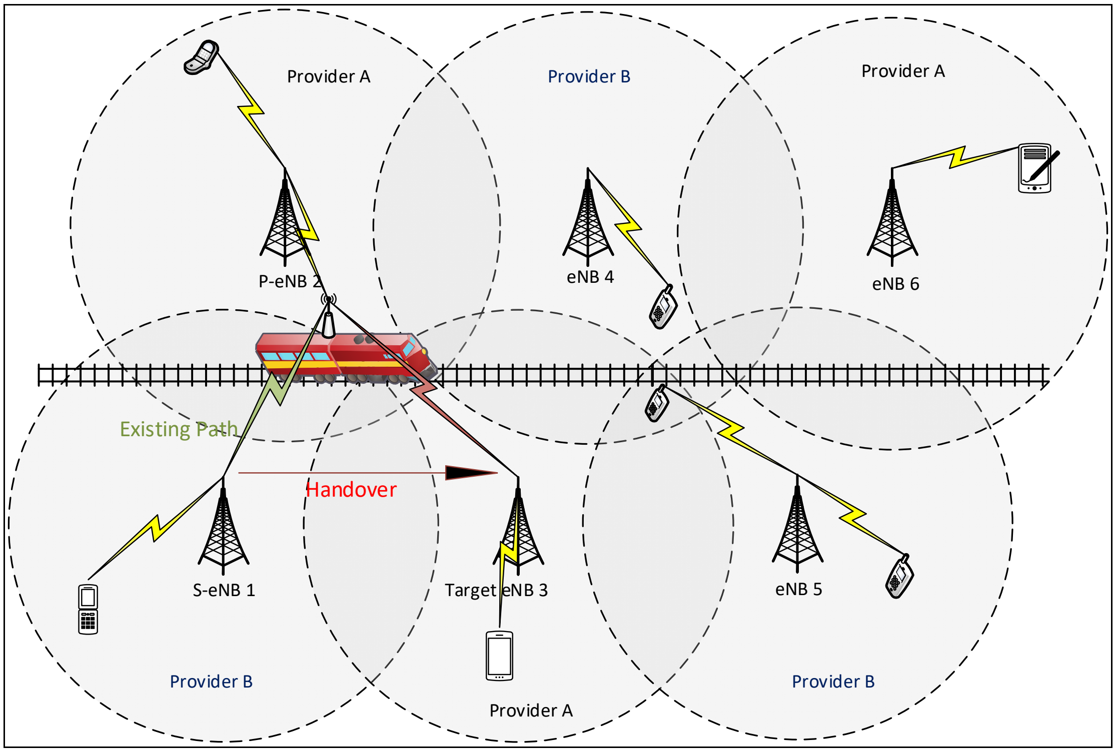



Figure 2.1 Illustration of LTE-A Dual-Backhaul links Mobile Relay in Railway Systems

## 2.3 Problem Formulation

In this section, the problem of implementing the carrier aggregation scheme via tow backhaul links will be formulated. The network is comprised of one mobile relay node with six eNBs. We assume that the six eNBs overlap with each other and belong to two different operators. The transmission range is assumed to be equal and denoted by R (eNB).

The SNR received by mobile relay node (MRN) from eNBs, denoted by $SNR_{eNB,MRN}$ , is given by:

$$SNR_{eNB,MRN} = \frac{G_{eNB,MRN}\, P_{eNB,MRN}}{P_N} \tag{2.1}$$

Where $P_N$ is the noise power at the relay level, $P_{eNB,MRN}$ is the power transmitted by eNB to Mobile Relay Node (MRN), $\mathrm{G_{eNB,MRN}}$ is the gain of the channel between eNB and MRN.

In terms of the bandwidth and signal to noise ratio, Shannon's theory gives an upper bound to the data rate. Using this theory, we can compute the data rate ($R_{eNB,MRN}$) as follows:

$$R_{eNB,MRN} = BW_{\ eNB,MRN} \log_2(1 + SNR_{eNB,MRN}) \tag{2.2}$$

Where $\mathrm{BW}_{\ \mathrm{eNB,MRN}}$ is the amount of bandwidth assigned to the MRN by each eNB. Hence, through the dual-backhaul links, the total data rate consists of the primary and secondary data rates as follows:

$$\begin{aligned} R_{Tot.} = {} & BW_P \log_2(1 + SNR_{P\_eNB,MRN}) \\ & + \ BW_S \log_2(1 + SNR_{S\_eNB,MRN}) \end{aligned} \tag{2.3}$$

Next, by considering the total calculated throughput and the required throughput, the packet loss of the system (PL) can be represented by:

$$PL = \begin{cases} 0 & for\ R_{Tot.} \geq R_{Req.} \\ 1 - \dfrac{R_{Tot.}}{R_{Req.}} & for\ R_{Tot.} < R_{Req.} \end{cases} \tag{2.4}$$

Where $R_{Tot.}$ is the total throughput achievable through carrier aggregation in the system and $R_{Req.}$ is the required throughput for the application.

During the handover procedure, the Blocking Probability rate (BP) of one call is calculated by the equation:

$$BP = \begin{cases} 0 & for\ R_{Tot.} \geq R_{Req.} \\ 1 & for\ R_{Tot.} < R_{min.} \\ 1 - \left(\frac{R_{Tot.} - R_{min.}}{R_{min.}}\right), & for\ R_{Req.} > R_{Tot.} \geq R_{min.} \end{cases} \tag{2.5}$$

Where $R_{min.}$ Represents the minimum acceptable throughput by UEs.

Our concern is to maximize the average throughput $R(i)$ for the user $(i)$ with the use of dual-backhaul links. Particularly, the objective function of the problem can be written as follows:

$$Max \sum_{i} R(i) \tag{2.6}$$

With the following constraints:

$$\begin{cases} R_{Tot.}(i) = R_P(i) + R_S(i), & \forall\, i \\ R_{Tot.}(i) \geq R_{min.}, & \forall i \end{cases} \tag{2.7}$$

Where $R_P$ and $R_S$ are the throughputs that are provided respectively by the primary and secondary backhaul links via the Carrier Aggregation technique.

### 2.4 Proposed carrier aggregation (Dual-Backhaul Links) scheme

In this chapter, Carrier Aggregation (CA) scheme is examined to improve the on-board user throughput as well as facilitating the group handover for such high number of user terminals. Conventionally, the carrier aggregation has only been introduced to ease the group handover. But none of the previous methods have considered the user throughput and the bandwidth offered by the eNB.

With the served on-ground UEs, the eNB of the LTE network may encounter some difficulties in providing the MRN of the train with the required capacity and satisfying the users with a high quality of service. Therefore, CA mechanism is employed to share the available bandwidth of two eNBs, through primary and secondary links, to fulfill the requirements of the moving passengers train. The MRN always keeps connected to at least two backhaul links. When RSS of eNB deteriorates, the mobile relay whose RSS goes down a certain threshold must handover to a target eNB.

This proposal gives many features such as: 1) It increases the user throughput by utilizing more than one backhaul link in the transmission especially in the urban regions; 2) Implementing CA mechanism with a minimum of a dual-backhaul links would make a major difference and facilitate the process of group handover, which is normally very hard and frequent due to the large number of users and the speed of the train.

### 2.5 Simulation setup, results and analysis

A scenario of six eNBs, a Mobile Relay Node mounted on a train, and up to 60 UEs, has been simulated to assess the performance of the proposed dual-Backhaul links scheme. Scenarios of traditional MRN and direct communication have been used to compare the performance of our strategy.

The model of propagation used in this simulation- between eNB and Carriage- is the path loss model for free space, which is given by:

$$Path\ Loss\ (PL) = 32.4 + 20log\ (f) + 20log\ (d) \tag{2.8}$$

Where f is the frequency (MHz) and d is the distance of MRN from the eNB.

For the case without MRN, every VUE directly communicates with the eNB and they will experience the penetration losses of train varying from 3.2 to 23.8 dB (Tanghe et al., 2008).

The simulation parameters of the LTE-A and Mobile Relay Node of the train are illustrated in Table 2.1. A comparison of a railway scenario with carrier aggregation, standard MRN, and without MRN has been given.

Table 2.1 Simulation Parameters

| **Parameter** | **Value** |
|---|---|
| Bandwidth of Channel (BW) | 20 MHz |
| Carrier Frequency | 2.2GHz |
| Transmission Power | 43 dBm |
| Transmission Gain | 20dB |
| Reception Gain | 2 dB |
| Supplementary Attenuation | 20 dB |
| Sensibility | -100 dBm |
| Cell Radius (R) | 700 m |
| Number of VUEs | 60 Users |

In the simulation, real-time traffic was used, as shown in the following table.

Table 2.2 Real-Time Traffic Parameters

| **Parameter** | **Value** |
|---|---|
| Traffic Flow | 1Mbps |
| Minimum acceptable Throughput | 500 Kbps |
| Acceptable maximum packet loss | 10 % |

MATLAB software was used to build and simulate our scenarios. We focus on the capacity of the system, group mobility and perform multiple series of comparisons between simulations. We are considering CA-MRN, conventional MRN and Free-MRN scenarios in LTE Railway networks.

Figure 2.2 illustrates the timeline for the average users' throughput against simulation time (81 Seconds).

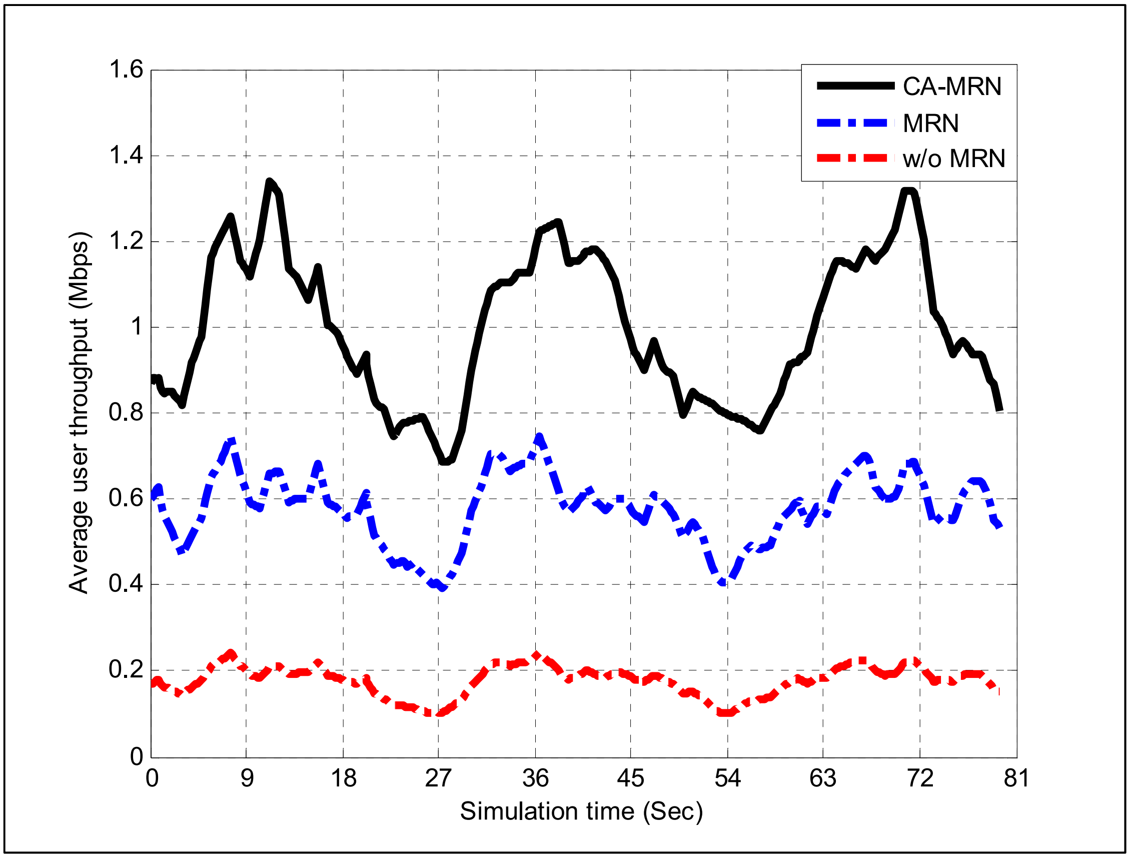



Figure 2.2 UE's Average Throughput

It can be seen in the figure that the CA-MRN mechanism performs better than the conventional MRN and w/o MRN schemes in terms of the average user throughput. The average user throughput of CA-MRN can reach 1.1 Mbps, while it barely reaches 0.6 Mbps when employing the standard MRN method. CA-MRN gives an 83.3% increase in the average vehicular users' throughput. This increase was obtained with the additional backhaul link which provides the transportation train with more bandwidth available. Moreover, for the case without MRN, the average user throughput falls severely to around 0.15 Mbps which is due to the VPL impact, in addition to the train path which is assumed to be at the cell edges.

Figure 2.3 demonstrates a comparison of packet loss performances between Mobile Relay System supported by CA-MRN scheme, a Mobile Relay System with just MRN and a scenario of Free-MRN scheme.

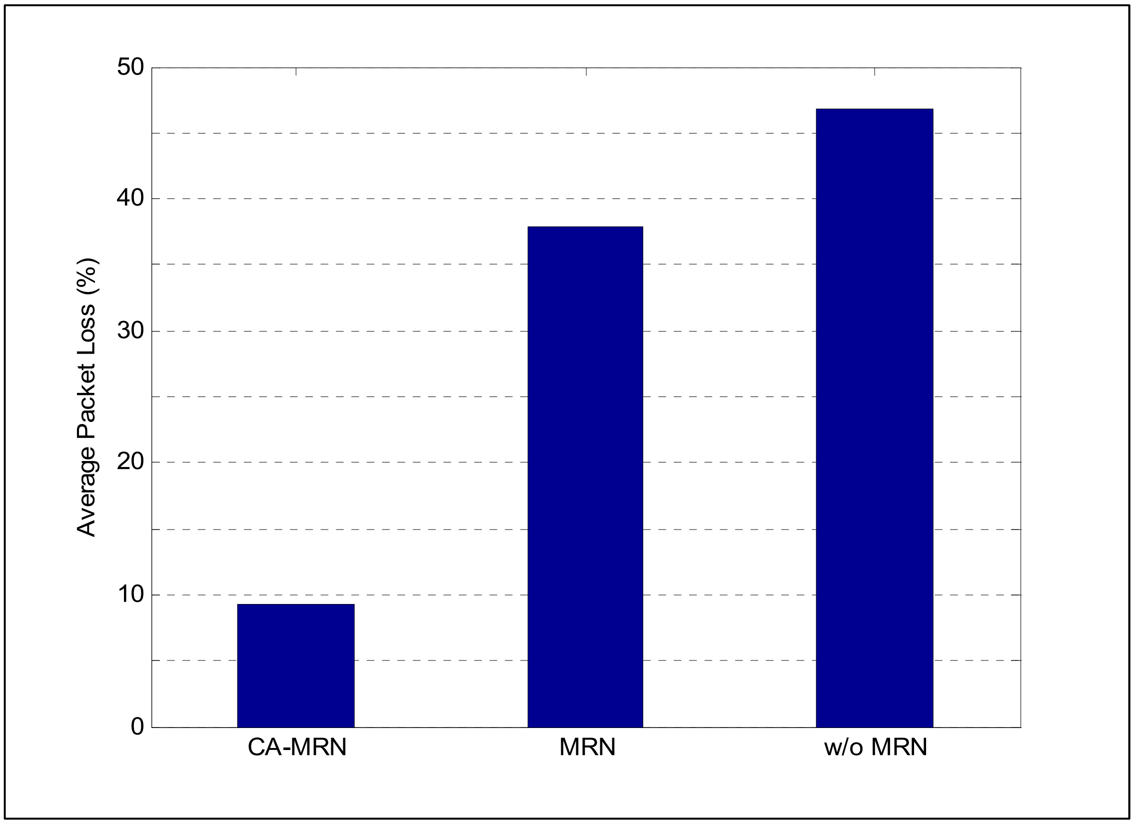



Figure 2.3 UE's Average Packet Loss Rate

The proposed solution proves that with employing CA-MRN for the HST, the VUEs can reach a very low average packet loss less than 10%, caused by the fact that most users are able to get the minimum required throughput. However, for scenarios with MRN without CA and without MRN, the average packet loss rate exceeds respectively 35% and 45% for all VUEs. These results are due to the low generated throughput.

Figure 2.4 shows the average blocking probability during the group handover process for our proposed scheme CA-MRN and the traditional MRN.

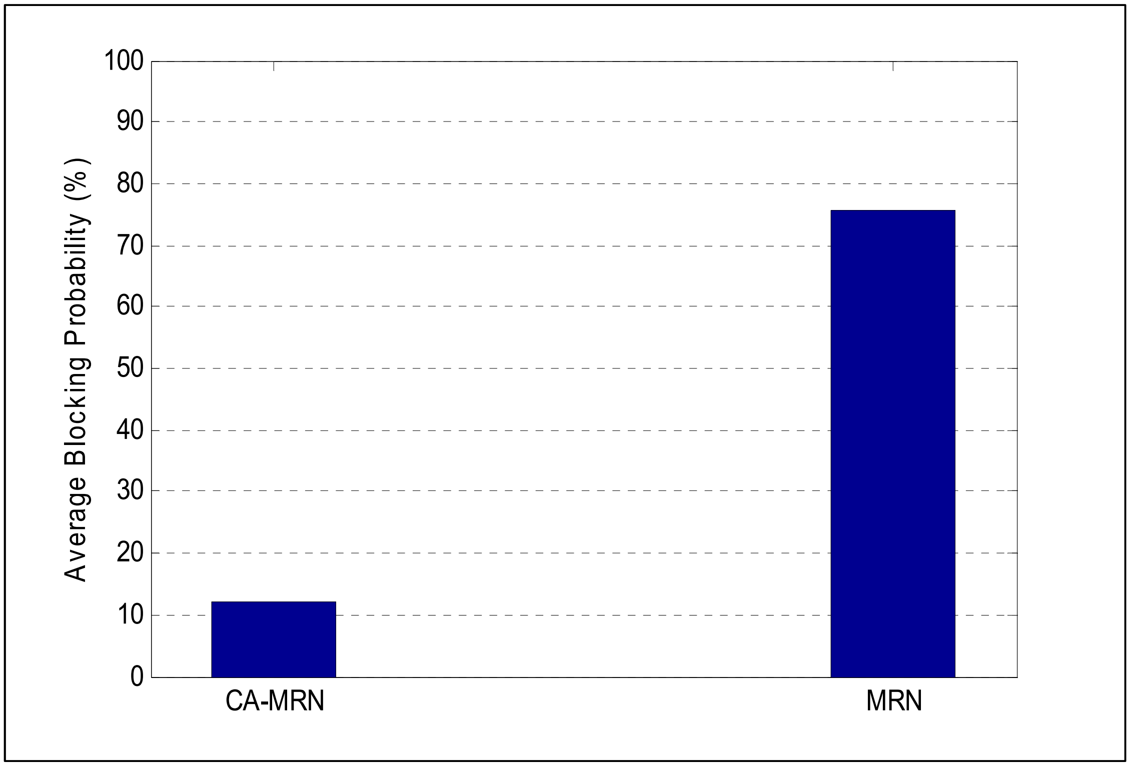



Figure 2.4 Average Blocking Probability Rate

It is noticed from the graph that each scheme has a different blocking probability rate. The proposed CA-MRN presents a remarkable low blocking probability rate over the MRN mechanism. In other words, the blocking probability rate reaches approximately 75% for the case of the MRN scheme, whereas it decreases to around 10% for the CA-MRN because more bandwidth is available by the proposed additional Backhaul Link during group handover.

The reason behind the degraded performance for the MRN scenario is that the target eNB cannot admit all the UEs during the traditional group Handovers due to bandwidth limitation. Previous results demonstrate that the performance of the CA-MRN is remarkably higher than that of the Free-CA MRN scheme in Railway scenarios. The simulation results show that our proposed scheme outperforms other schemes proposed in the literature because of the cooperation between different operators in providing dual-backhaul links and thus more bandwidth to the Mobile Relay Node.

## 2.6 Conclusion

Enhanced throughput and guaranteed coverage extension can be provided by using the mobile relaying techniques in LTE-A system Networks. 3GPP started supporting mobile relay in LTE-Networks (Rel.12). In this section, we presented the carrier aggregation-supported mobile relay for railway LTE-A networks. The performance of the novel scheme was evaluated in terms of average user throughput, packet loss and outage probability. Results show that better performance can be accomplished by applying dual-backhaul links with carrier aggregation scheme. In next chapter, we will investigate the UE-eNB association to minimize the number of blocked vehicular users in the network when the network resources are scarce and enhance the QoE of vehicular users.

# CHAPTER 3

# MATCHING GAME-BASED USER ASSOCIATION APPROACH FOR LTE-A MOBILE RELAY SYSTEMS

## 3.1 Introduction

Relay-enhanced LTE is a many-to-one association between Relay Nodes (RNs) and eNB (i.e., several relay nodes can be connected to one eNB but an RN is connected only to one eNB) (Teyeb et al., 2009). Though that is a simple solution to enable relaying in LTE, the end-to-end performance of relayed UEs will be constrained by the capacity available on the backbone link that is accessible through the $S1$ interface. Radio resources might be sufficient at the relay link but congestion in the backbone (i.e, on the $S1$ link) can degrade the performance of relayed UEs. Nevertheless, we may have unfavourable conditions or insufficient radio resources on the relay link (i.e, $Un$ link) with the DeNB despite there are lightly loaded neighbouring cells. Hence, many-to-many connections between RNs and eNBs through the $Un$ interface are introduced to use the sum of the available capacity in all the neighbouring cells. User association is when the user or relay node is assigned to different stations in the system (Trabelsi et al., 2017). Conventionally, users or relay nodes are assigned to only one eNB at a time. It is proved that the achievable throughput of any terminal depends on the users-eNBs assignment decisions. In this work, for a distributed relaying system, a user association method that increases the number of users admitted in the system will be developed (i.e. minimizing the blocking rate at any given time). In addition, with the mobility of the relays considered, the handover failure rate will greatly be minimized. Specifically, this chapter focuses on the first critical problem aforementioned that considerably determines the possibility of admitting a call in the network (Macro cell) without affecting the quality of service required by the call context. A call is rejected if the above condition is not satisfied. The admission control (AC) establishes the rules of acceptance of different call types (new or handover) in the network to ensure a very low block rate and an optimal call acceptance. The challenges in the user association are to minimize the number of blocked vehicular users in the network when the network resources are scarce and enhance the QoE of vehicular users. In our work, we propose a user association algorithm called the Chance-Based Deferred Acceptance Matching (CBDAM) Algorithm.

### 3.2 System Model

The target system is an Intelligent Public Transportation System environment, represented by a large-scale Evolved Universal Terrestrial Radio Access (E-UTRAN). The studied model is derived from similar models used in our past work (Addali et al., 2016). In this architecture, the Radio Access Network (RAN) consists of I eNBs entirely covering the geographical area overlaid with three MRNs. MRN can have multiple parents for their back-haul link connection.

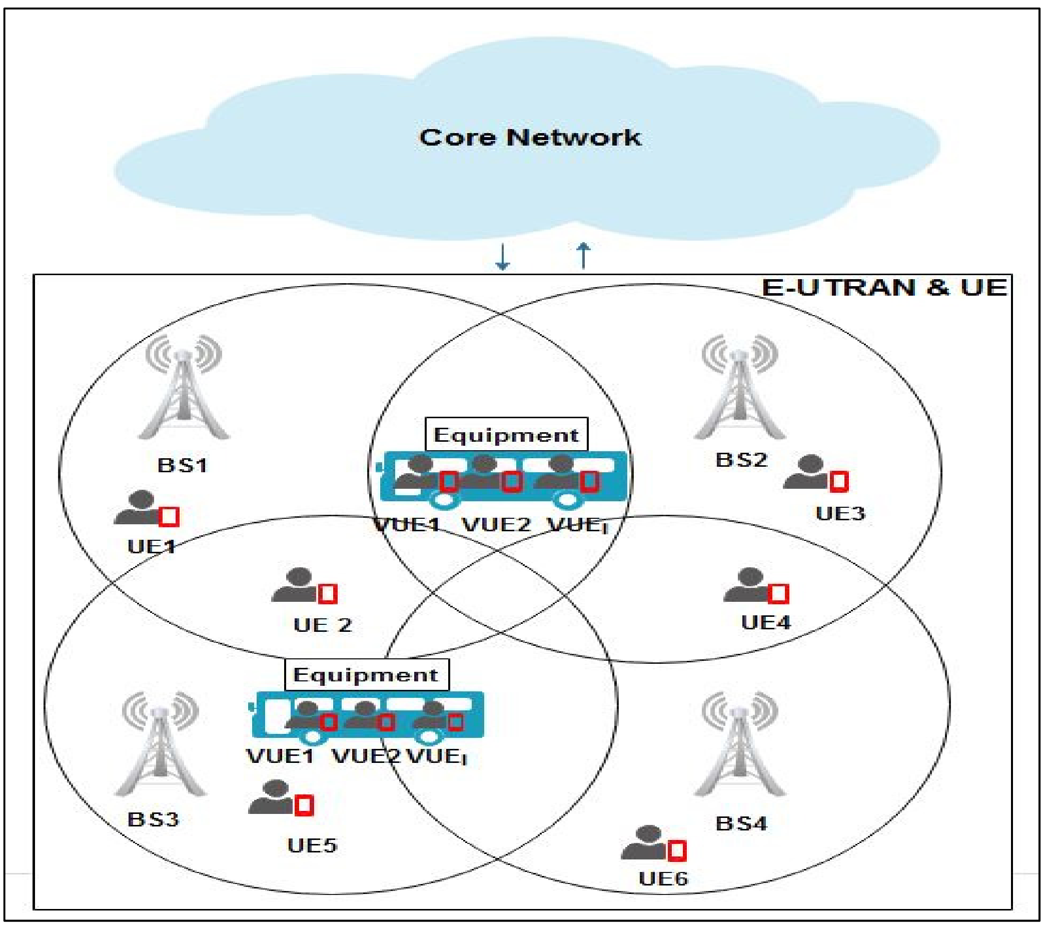



Figure 3.1 LTE-A Mobile Relay System Description

The target system model is depicted in Figure 3.1 which is a modified version of the model described in our previous work. Each eNB serves J mobile macro users and mobile relay nodes through LTE wireless links. We assume that MRN is a heterogeneous relay that uses Wi-Fi on the access link while using the LTE network on the back-haul link. The macro users are uniformly distributed and assumed static in this study, whereas vehicular users are assumed to be at the same location as their MRN. In summary, our system consists of the following components:

- Base stations (eNBs).
- Mobile Relay Equipment: this component is responsible for serving vehicular users by associating with one or more base stations (eNBs) according to their needs.
- Users: two types of users are considered; macro and vehicular users.

The system total bandwidth is split into three orthogonal sub-channels for system downlink transmissions. Thus, we set the frequency reuse factor for this model to 1/3, such that the signal to interference plus noise ratio will only exist between MRN and macro users, and vehicular users will not be affected by any sort of interference.

### 3.2.1 System Parameters

The signal from an eNB towards the mobile relay node (MRN) or macro user equipment (MUE) $(i)$ is assumed to experience a long-distance path-loss law $pl_i(d)$ with exponent = 2 (i.e. free space propagation model). The received power from $eNB_j$ at an MUE or MRN $(i)$ at relative position d, d is in meters, is expressed as.

$$P_{j,i}^{R_x} = P_{j,i}^{T_x}\, pl_i(d) \tag{3.1}$$

Where $P_{j,i}^{T_x}$ denotes the transmit power of $eNB_j$ . In our work, we intend to apply the rural path loss model of the technical report (TR36.942) specification which represents the typical propagation environment for a bus travelling through the countryside, which is known as Hata model and expressed by:

$$pl_i(d) = 69.55 + 26.16\ \log_{10}(f) - 13.82\ \log_{10}(Hb) + [44.9 - 6.55 \log_{10}(Hb)] \log(d) - 7.78(\log_{10}(f))^2 + 18.33\ \log_{10}(f) - 40.94 \tag{3.2}$$

Where: $d$ is the eNB-UE separation in meters, $f$ is the carrier frequency in$MHz$, and $Hb$ is the eNB antenna height above ground in meters.

The $SINR_{j,i}$ received by MUE or MRN $(i)$ from $eNB_j$ is given by the equation:

$$SINR_{j,i} = \frac{P_{j,i}^{R_x}}{P_{k,i}^{R_x} + N} \tag{3.3}$$

Where $N$ is the noise power at the terminal level, $P_{k,i}^{R_x}$ is the total power received by MUE or MRN $(i)$ from $eNB_k$ (i.e. interfering eNBs), and$P_{j,i}^{R_x}$, is the power received by MUE or MRN $(i)$ from $eNB_j$ (i.e. desired eNBs). No penetration loss is considered for the MRN setup and the MRN is modelled as a super active UE representing a set of vehicular users (VUE).

With that $SINR_{j,i}$, Shannon's formula determines the theoretically possible throughput, $R_{j,i}$, achievable by MUE or MRN $(i)$ from $eNBj$ as follows:

$$R_{j,i} = \gamma_{j,i} \log_2(1 + SINR_{j,i}) \tag{3.4}$$

Where: $\gamma_{j,i}$ is the channel bandwidth allocated by $eNB_j$to MUE or MRN $(i)$ in Hz, and $SINR_{j,i}$ is the signal to interference plus noise ratio at the receiver $(i)$.

Bandwidth resources allocated by eNBs to every user are characterized by the access techniques adopted by the network. Specifically, LTE-Advanced wireless networks use two types of multiple access techniques: Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink, and the Single-Carrier Division Multiple Access (SC-FDMA) in the uplink. OFDM performs greatly in frequency selective fading channels and combats their most common result inter-symbol interference (ISI). Moreover, it enhances the spectral efficiency by not considering band guards and provides multiple bandwidth management and scheduling frequency domain. In fact, OFDM is a special case of the frequency division multiplexing (FDD) in that it utilizes multiple carriers with the simultaneous transmission. Subcarriers spacing (or symbol rate) plays a very significant role in making the adjacent sub-carriers cancel one another. Subcarriers are spaced with 15 KHz in LTE-A no matter the transmission bandwidth.

LTE-A allocates and schedules resources dynamically in the frequency domain based on Resource Blocks (RB). Every consecutive 12 sub-carriers form one RB that has 180 kHz bandwidth in the frequency domain and 6 or 7 OFDM symbols (i.e. 1 slot) in the time domain. A resource block (RB) is the smallest unit of resources that can be allocated to a user. However, despite the existing scheduling schemes, only one RB is assumed per user for the assignment problem.

### 3.2.2 Problem Formulation

This is a resource-constrained matching problem that is restricted by the users' resource requirements and eNB capacity limits. Essentially, the best matching problems are integer or mixed-integer programs. We have formulated this problem as a many-to-one matching process, in which each user $i \in I$ requires a certain amount of resources from some eNB $j \in J$ that has limited resources. The problem is represented by the notion $M : 1, RC,$ and $NA$ which stands for many to one matching process, resource constraint, and the number of the admitted users, respectively. Our objective is to maximize the number of mobile relay users admitted in the system as well as the total system performance. The first constraint ensures the minimum acceptable throughput by every user $R_{min}$, while the second constraint prevents the process from exceeding the eNB capacity $C_j$. The third constraint ensures that every user can only be associated with one eNB. This notion was introduced and explained in (Moghaddam et al., 2017). The many-to-one matching is formulated as follows:

$$M: 1/RC/+NA \tag{3.5}$$

$$Maximize \sum_j \sum_i x_{j,i} \tag{3.6}$$

$$Subject\ to \sum_i R_{j,i} \geq R_{min}, \forall i \in I, \tag{3.7}$$

$$\sum_i \gamma_{j,i}\, x_{j,i} \leq C_j, \forall j \in J, \tag{3.8}$$

$$x_{j,i} \in 0,1, \qquad \forall i \in I, \qquad \forall j \in J, \tag{3.9}$$

$$\sum_{j \in J} x_{j,i} = 1, \forall i \in I,$$

Where: $\{x_{j,i} | j = 1,2, \dots M\}$ is the best matching decision variable

$$x_{j,i} = \begin{cases} 1\ if\ j\ and\ i\ are\ matched \\ 0\ otherwise \end{cases}$$

Unlike static matching models, and due to the mobility of the RNs, the time dimension is incorporated into the system state. Furthermore, the dynamic matching model considers the real-time variation of the wireless channel. However, at a certain moment, the matching is performed as a static model.

### 3.2.3 Preference Ranking Criteria

Users rank their potential access point (AP) over another based on the benefit (utility) they might achieve from APs:

$$u_i = \ln(1 + R_{j,i}) \tag{3.10}$$

Where: $u_i$ is the user's utility function, and $R_{j,i}$ is the rate achievable by user $i$ from $eNB_j$. It is important to mention that MRN will rank eNBs on behalf of its vehicular users. Moreover, MRN can rank eNBs according to the number of back-haul links supported by its equipment. So, in our study, since MRN supports dual back-haul links, MRN will select two eNBs with the highest SINR. For the eNB preference ranking, on one hand, eNBs prefer to maximize their own capacity by associating with users with high capacity and it is calculated by summing the users' individual utilities:

$$S_j = \sum_i u_i \, x_{j,i} \tag{3.11}$$

Where: $S_j$ is the $eNB_j$ utility, and $x_{j,i}$ is the association indicator ($x_{j,i}$= 1 if user $i$ is associated with $eNB_j$ , $x_{j,i} = 0$ otherwise).

On the other hand, they might rank based on the chances that the users can find. For instance, if some user has only one eNB option and another one has two options, the eNB will rank the former as its most preferred user while the latter as its second preferred user. Both criteria we have used in our work in different ways.

## 3.3 Chance-Based Deferred Acceptance Matching Algorithm

In relaying systems, the terminal association or admission control should be handled between eNBs and its subordinate relays or users. However, that would generate a lot of delays and high control overhead for the Relayed UEs, while the resource availability of the relay cannot be considered as an accurate metric for making the admission decisions. Thus, the resources on the wireless back-haul link must be the main parameter to guarantee the availability of resources through the wireless path. The wireless backhaul link is the only path for transmitting the vehicular users' traffic, so whatever throughput vehicular user's requests, BSs will decide and allocate according to its resource management policy. Access link throughput is limited by the resources assigned to the back-haul link. Therefore, we employ the matching game theory to solve the MRN association problem with eNBs. We are using a modified deferred acceptance algorithm to perform the matching negotiation between eNBs and relays, as well as users. It is a matching between two sets (eNBs and users) with the elements of users' set proposing and the elements of eNBs set responding. This algorithm guarantees stable matching in polynomial time and does not necessarily yield the optimal matching with respect to the mutual preferences of the sets' elements.

First, eNBs and MRNs or users build their preference matrices: $P_{ji} = (\text{a};\ \text{b})$ where: $a$ is the normalized preference of $i$ eNB for $j$ users. $b$ is the normalized preference of $j$ users for $i$ eNB. $a$ and $b$ can be formalized to be the preference degree (e.g a is set between 1 and 4 means 1: least preferred and 4: most preferred).

Initially, eNBs and UEs are assumed to be free. Then, every user and MRN propose to its most preferred eNB. Respecting their limited quota, each eNB creates a waiting list, holds the most preferred UEs and rejects the rest. Then, we update both the preference lists of the eNBs and UEs after every iteration, such that users remove eNBs that reject their previous applications and rank again while eNBs re-update their lists according to the chances left for the UEs. However, when an eNB receives proposals that are more than its capacity, eNB holds the most preferred ones and rejects the others.

The waiting list is updated out of the top proposers (but not exceeding its quota). All the algorithm steps are illustrated in the pseudo-code.

Algorithm 3.1 Chance-Based Deferred Acceptance Matching (CBDAM) Algorithm

**Require:** eNBs Capacity and Locations, MRN (VUEs information) and Macro users' location information, SINR, and MRN speed.
**Ensure:** Converge to a stable matching$\mu$.
**Procedure:** CONSTRUCT PREFERENCEMATRIX ($P_{ij}$)
With entries $Pij = (pi \rightarrow j, pj \rightarrow i)$.
$pi \rightarrow j$ denotes the normalized preference list i for $j$.
$p_{j\rightarrow i}$ denotes the normalized preference list j for $i$.
Users construct their $pi \rightarrow j$ based on achievable throughput.
eNBs construct their $pj \rightarrow i$ based on their Capacity and users' chances $pi \rightarrow j$.
**end procedure**
**Phase 1 - MRN and Macro Users propose to their most preferred eNBs (maximum $\boldsymbol{pi \rightarrow j}$)**
**for all** eNB ($j$) **do**
    **for** each user ($i$) **do**
        **if** Capacity of BS ($j$) ≤ maximum capacity and user ($i$) has been ranked first by the BS ($j$) **then**
        **else**
            BS ($j$) rejects user ($i$)
        **end if**
    **end for**
**end for**
**Phase 2 - Update eNBs ($\boldsymbol{pj \rightarrow i}$) and users ($\boldsymbol{pi \rightarrow j}$)**
Repeat
eNBs re-rank all UEs rejected in the previous step based on the chances left.
UEs and MRNs remove all the eNBs that reject proposals and update their preference matrices.
**for all** eNB ($i$) **do**
    **for all** rejected users ($j$) **do**
        **if** Capacity of BS ($j$) ≤ maximum capacity and user ($i$) has been ranked by BS ($j$) **then**
            BS ($j$) holds user ($i$)
        **else**
            BS ($j$) rejects user ($i$)
        **end if**
    **end for**
**end for**
Until no matching changes in the previous iteration.

### 3.4 Simulation Setup

We used the MATLAB platform to simulate our proposed solution developed in the previous sections. We first validate the simulation model and show some advantages of employing matching game theory to solve the downlink user association problem in multi-backhaul mobile relay architecture, such as the number of users admitted and their total throughput. We compare the proposed approach performance to the base approach which is known as the received-signal-strength (RSS) scheme. Then, we evaluate the performance of the proposed approach in a sub-urban environment scenario composed of four cell eNBs overlaid with three mobile relay nodes. Figure 3.2 illustrates the network simulation model.

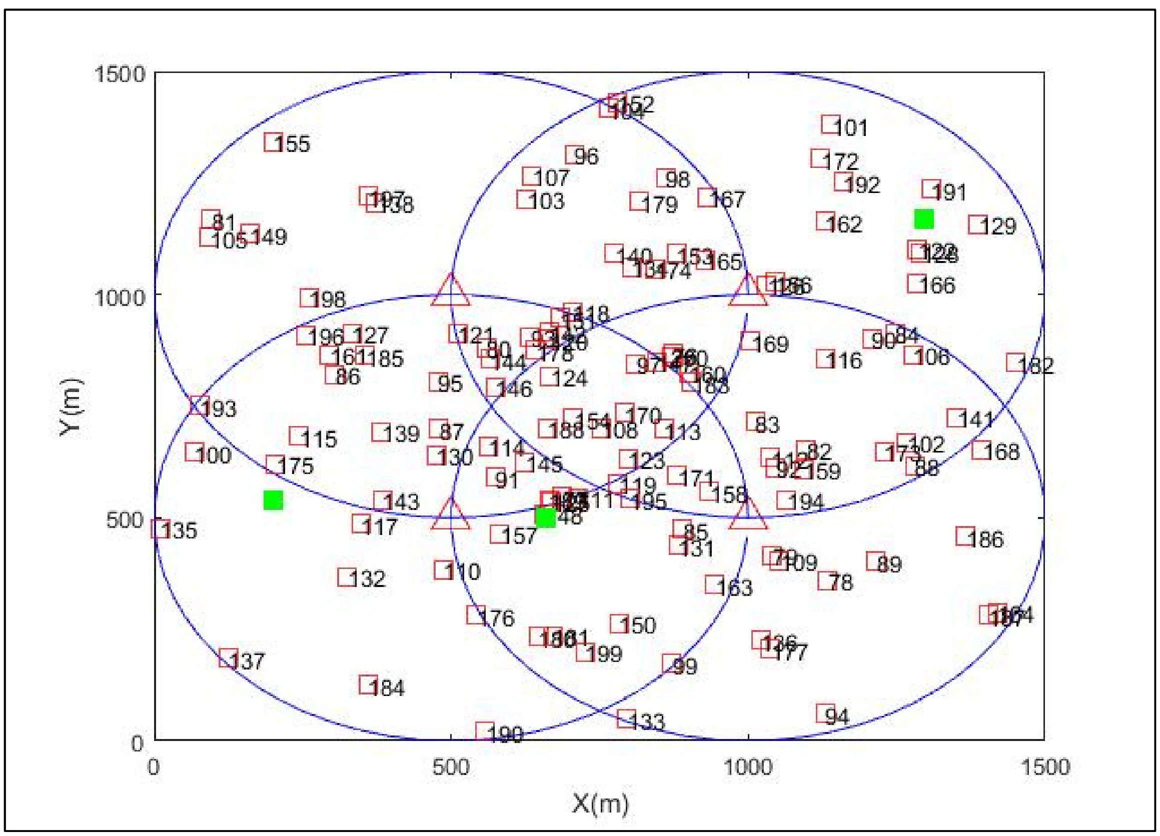



Figure 3.2 Network Simulation Model

The Network simulation parameters for the heterogeneous network are presented in Table 3.1. The simulation results of the total system throughput, the number of users admitted in the system, the blocking probability of vehicular users, and the vehicular user's handover failure rate, are collected to plot the figures.

Table 3.1 Simulation Parameters

| Parameter | Value |
|---|---|
| $P_{BS}$ | 43 dBm |
| $C$ | 10 MHz (BS) |
| $P_{thr}$ | -100 dBm |
| $r_{min}$ | 1 Mbps |
| $V$ | 60 Km/hour |
| $N_s$ | 7(4BSs + 3 MRNS) |
| $M$ | 200 UEs (155 Macro UEs +45 Vehicular UEs) |

### 3.5 Simulation Results

The proposed solution uses matching game theory to solve the user association problem when a macro user or a mobile relay has several stations in their range from which to choose. Matching game theory allows users to declare their preferences for association with eNBs instead of the usual metric RSS. We selected three performance metrics to compare our proposed algorithm with the traditional algorithm, which are the blocking probability for the system and vehicular users, system throughput, and handover failure rate for vehicular users.

Figure 3.3 shows the total number of users admitted to the system for CBDAM and RSS algorithms. From these results, it can be noticed that under the proposed algorithm, the number of UEs admitted by the system increases. This performance is a result of resource availability at eNBs, algorithm capability to shift users to neighbouring lightly-loaded cells and user (or MRN) locations.

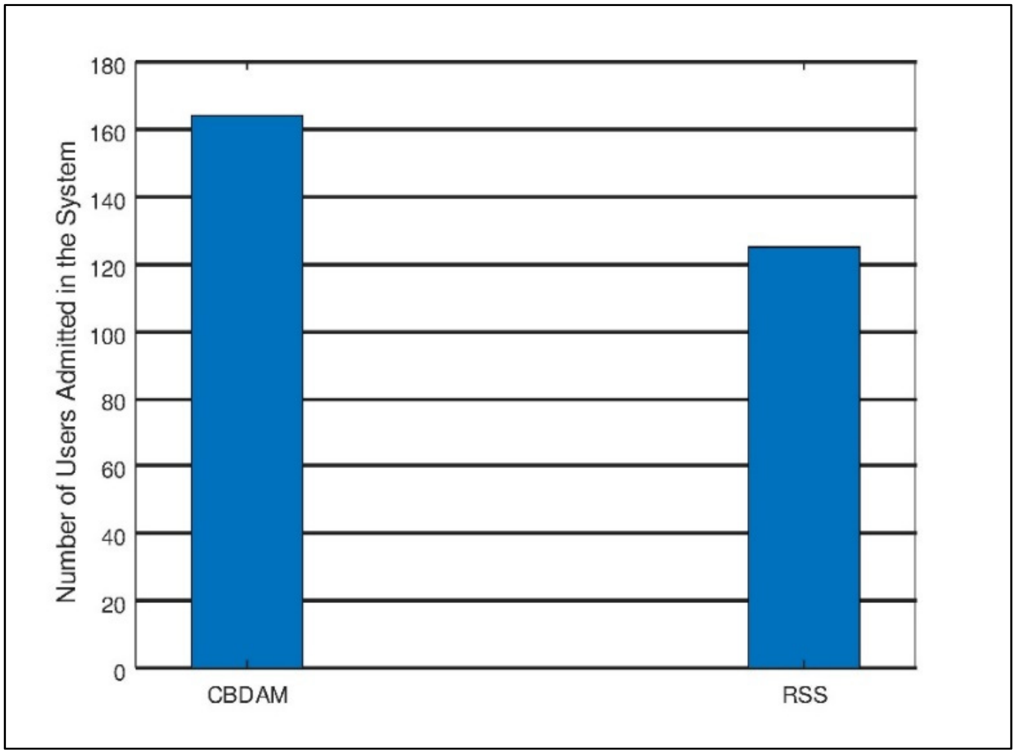



Figure 3.3 Number of users admitted in the system

Figure 3.4 describes the total system throughput over the simulation time. The proposed algorithm shows higher throughputs under a congested network. This performance can be explained by the matching theory algorithm capability in shifting users from the loaded eNBs to lightly loaded adjacent eNBs. Throughput fluctuations are because of the path that MRNs take (i.e., MRN directions might be towards or outwards the eNBs centres).

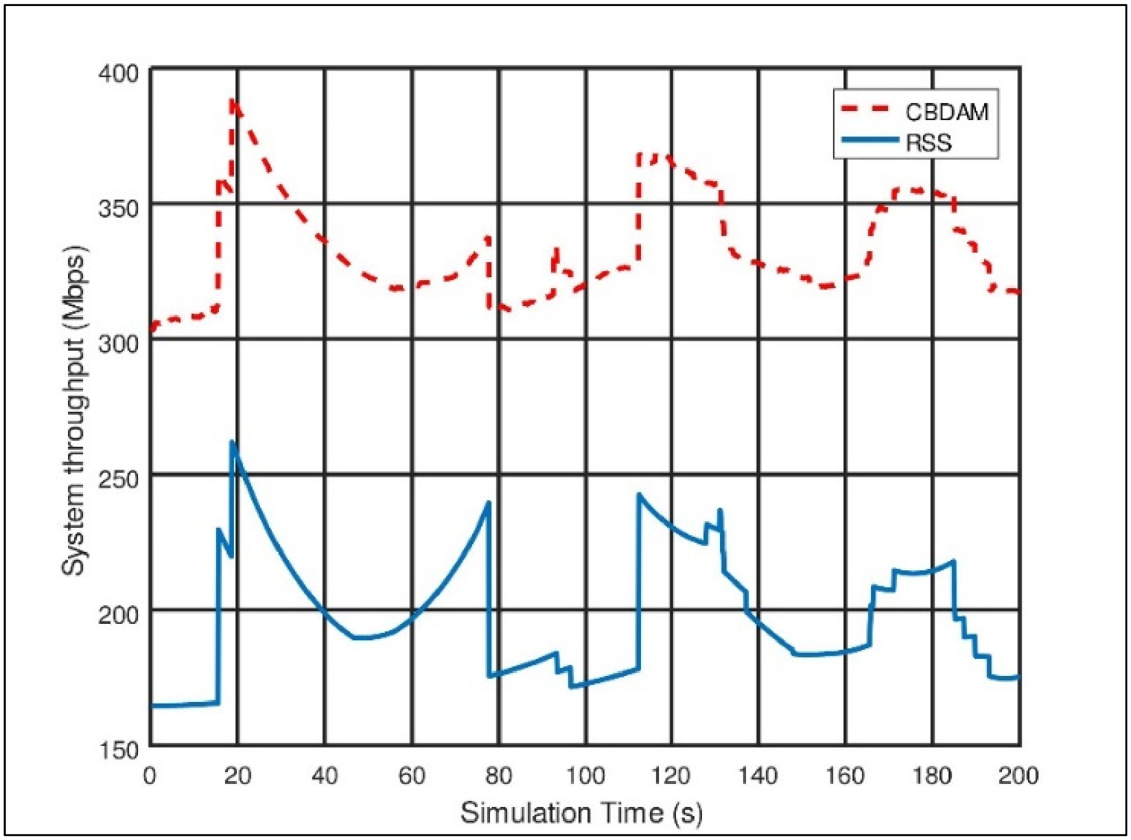



Figure 3.4 Total System Achieved Throughput

Since we mainly introduced this scheme to accommodate vehicular users, we show the blocking probability for the mobile relay users in Figure 3.5.

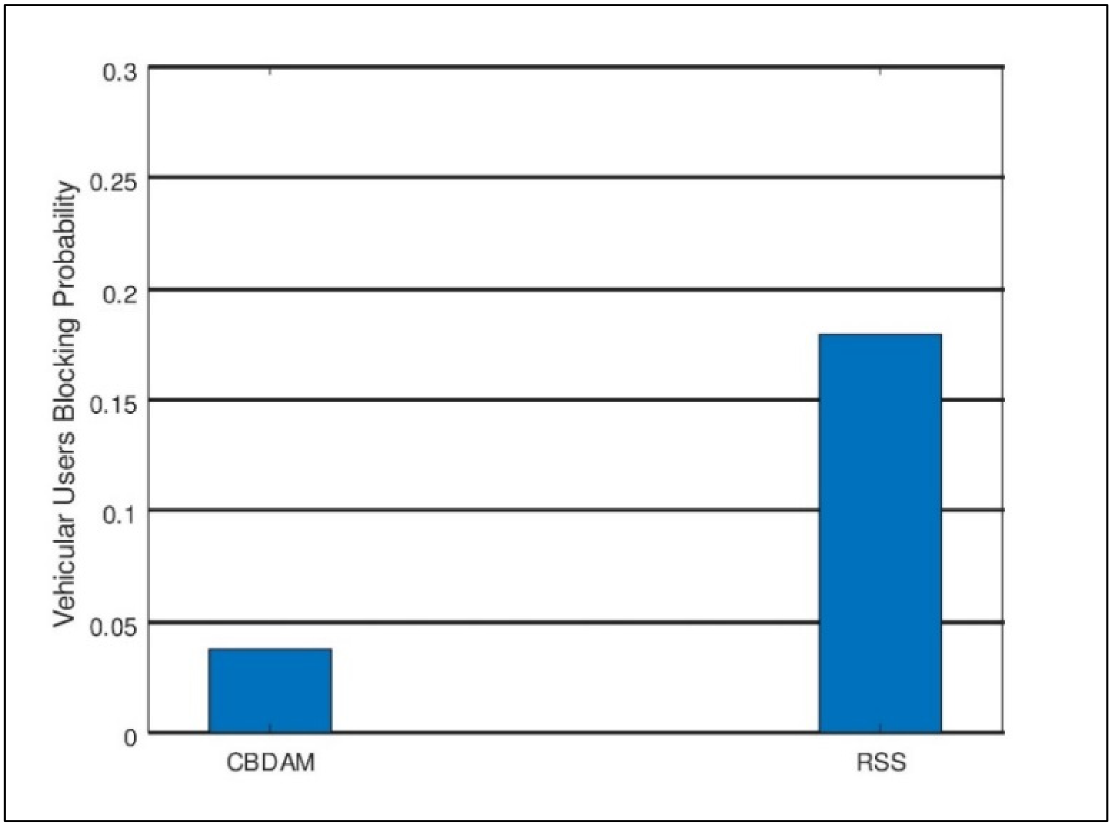



Figure 3.5 Vehicular Users Blocking Probability

The algorithm demonstrates outstanding performance when employed to address the group handover for MRN users. The received signal strength threshold to trigger the handover is set to be equal or less than - 80dBm. We assume that all the vehicular users are active during the handover process. Figure 3.6 illustrates that our proposed algorithm outperforms the RSS scheme, that is, the handover failure rate has decreased by 27%. This performance is a result of the negotiation that the matching algorithm performs between the sets of the network elements. Generally, matching algorithms enhance the performance of the system with respect to the given criteria by sub-optimally matching users to base stations. In contrary to the centralized optimization algorithms in the literature, the matching game introduced a near-optimal solution because it shows a better performance under similar Qos and capacity limitation constraints.

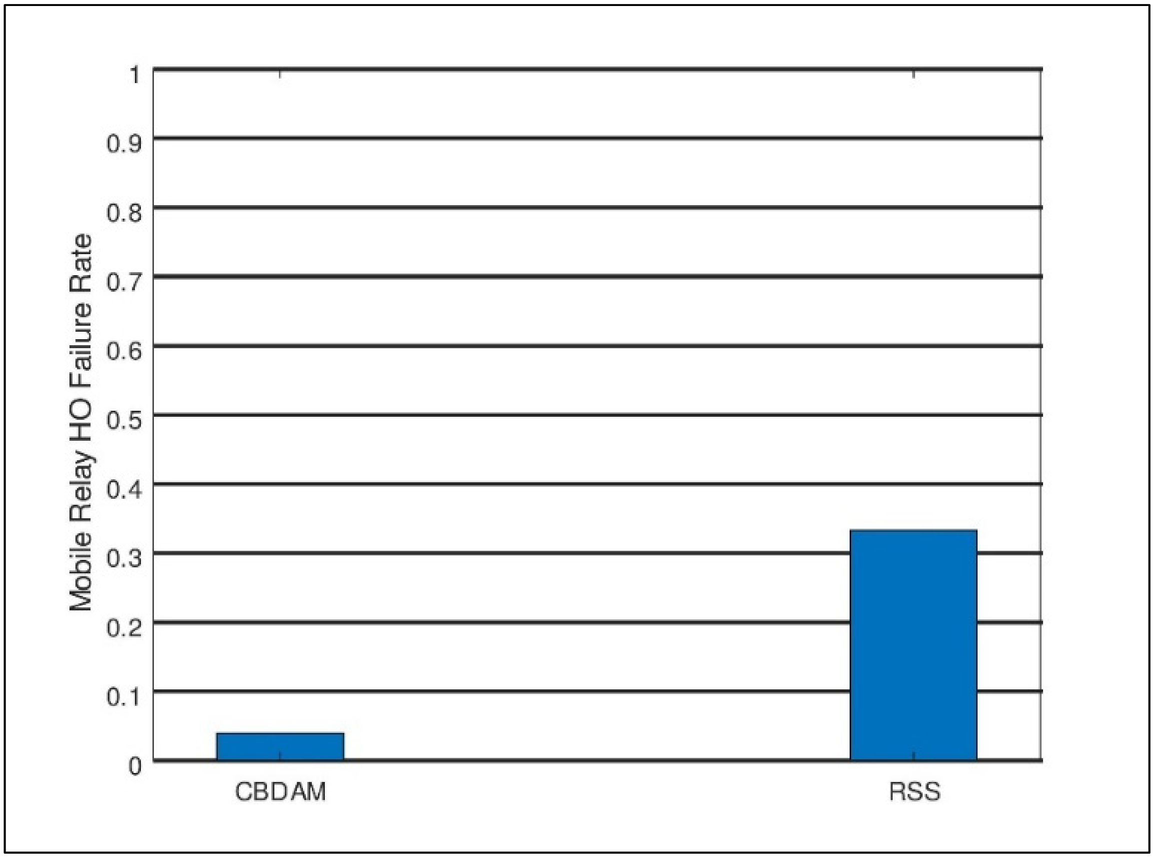



Figure 3.6 Mobile Relay Handover Failure Rate

## 3.6 Conclusion

This work investigates the user association problem in the downlink LTA-A network with mobile relay nodes.m . The challenges in the UE-eNB association are to minimize the number of blocked vehicular users in the network when the network resources are scarce and enhance the QoE of vehicular users. However, this is a complicated process for large-scale heterogeneous networks. Thus, a matching game-based user association scheme was introduced for solving the assignment problem to increase the number of admitted users in the system under resource constraints and QoS requirements. The designed algorithm is called the Chance-Based Deferred Acceptance Matching (CBDAM) Algorithm.

Particularly, and as illustrated by results, the new proposed approach maximizes the number of users admitted in the system and the total system throughput by assigning each user (macro or vehicular) to the stations based on negotiations between eNBs and UEs. With mobility existence, the handover failure rate is significantly decreased for the VUEs. In the next chapter, we will study the load imbalance across a small cell network due to the small area of small cells and the non-uniform distribution of UEs. We will consider both the operator and UE utilities in the load balancing process. Also, the offloading order will be investigated so that each overloaded small cell get the chance to offload.

# CHAPTER 4

# DYNAMIC MOBILITY LOAD BALANCING FOR 5G SMALL CELL NETWORKS BASED ON UTILITY FUNCTIONS

## 4.1 Introduction

Mobility of UEs in a small cell network with low service area cells may cause load-imbalance across the cells in the network. The performance of the network in terms of capacity and handover success rate degrades as a result of such an unbalanced load. The shortage of resources in the overloaded small cells leads to poor QoS and increases the handover failure rate when UEs intend to enter those cells though they have lightly loaded neighbouring cells. Consequently, resources of the unloaded cells remain unutilized though some overloaded neighbouring cells cannot meet the QoS requirements. Thus, the network needs proper configuration and management mechanisms such that the QoS is improved.

System parameters are adjusted manually in the existing networks to reach high levels of operational performance. However, such manual tuning is becoming difficult with the fast evolution of networks. Self-organized network (SON) was introduced to configure, optimize, and heal itself automatically in LTE, and hence decrease the operational complexity (3GPP, 2014). SON algorithms are categorized into three classes: centralized, distributed, and hybrid. SON has several components such as mobility load balancing (MLB), frequent handover mitigation (FHM), mobility robustness optimization (MRO), and interference management (IM), that help small cells to deliver carrier-grade performance. MLB distributes the UEs load among small cells to enhance the QoS and to increase system capacity. MLB utilizes cells load information to optimize the cell boundaries to offload UEs. SON uses mobility/handover parameters for load balancing.

MLB distributes the load among the small cells by adjusting the mobility parameters (i.e., handover parameters) according to their load statuses. To shift the candidate UEs, the cell individual offsets (CIO) of the serving and neighbouring cells are adjusted by UEs based on the reported measurements. However, improper handover decisions and offloading sequence for overloaded cells in MLB might cause an inefficient usage of resources or degrade the service. For example, as shown in Figure 4.1, small cell A is under-loaded with a load of 50%, whereas small cells B, C, and D are overloaded and have load values of 70%, 75%, and 80%, respectively. If the fixed MLB algorithm is adopted, they sequentially take the highly loaded cells from the list in the order of cell load and offload their excessive load to the lightly neighbouring cells. As a result, no cell can unload, but cell B, which comes third in the order and has an under-loaded neighbour cell A.

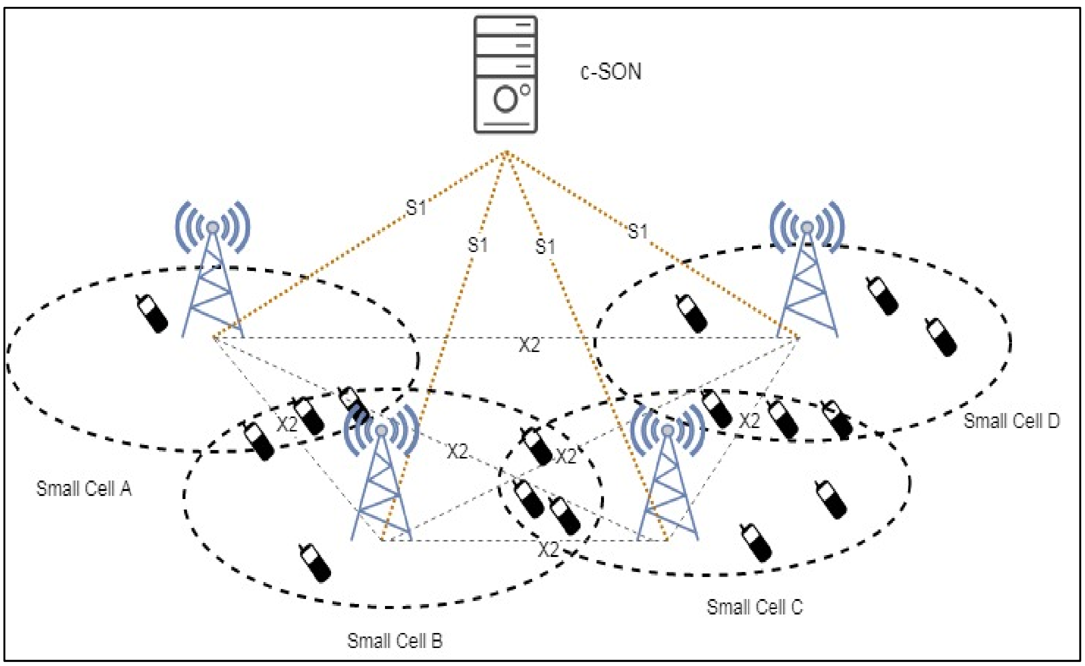



Figure 4.1 Radio Access Network Architecture with a c-SON controller

In this thesis, we introduced a Utility-based Mobility Load Balancing algorithm (UMLB) by considering both the operator utility and the user utility at the same time for each lightly loaded neighbouring cell. The main reason for choosing the utility function is that it expresses the satisfaction of any metric with a numerical value (e.g., cell load, reference signal received power (RSRP), or reference signal received quality (RSRQ)). Therefore, it is easier and less complex to compare these numerical values together to obtain the best decision during a handover process compared to the heuristic algorithms used in (Fedrizzi et al., 2016). The algorithm starts by determining the edge-UEs of an overloaded cell that needs to be offloaded to lightly-loaded neighbouring cells. The offloading process takes place by handing over each candidate of edge-UEs of the overloaded cell to the best neighbouring cell by calculating the aggregated utility for each neighbouring cell. The aggregated utility is a function in the operator utility and the user utility. The operator utility is calculated for each potential handover based on the load of the neighbouring small cells.

On the other hand, the user utility calculation is based on the sigmoid function by considering different criteria (e.g., delay, data rate, etc.) for each edge-UE involving a handover process. Furthermore, we introduced a new term named load balancing efficiency factor (LBEF) that considers a load of neighbouring cells and the edge-UEs for each overloaded cell. This factor specifies the sequence of overloaded cells for the UMLB algorithm operation.

## 4.2 System Model

This section defines the network model that will be investigated throughout the chapter. Next, the most important system constraints are determined. Finally, we explain how the cell load is represented and calculated.

### 4.2.1 Network Model

First, we introduce the network model, its parameters and assumptions. In this work, we investigate a homogeneous network of several small cells indicated by the set $S = \{1,2 \ldots S\}$, as depicted in Figure 4.1. The small cells belong to the same operator and operate in an open access mode. The centralized SON (c-SON) is adopted here. In c-SON, some optimization functions are executed at the Operation and Management system (OAM), while others are executed at eNBs.

The small cell network adopts two types of interfaces, as shown in Figure 4.1. Small cells connect to c-SON via the S1 interface (Yamamoto et al., 2012). However, they communicate with each other over the X2 interface. Small cells share handover-related information via X2 interface to execute handover of UEs from cell to another. Not only handover management is performed over X2 interface, but also load management such as resource status, and traffic load can be provided over the X2 interface (3GPP, 2019). The c-SON at the OAM periodically gathers information form small cells' c-SONs and uses them, if any overloaded cell is detected, to optimize and update the small cells' handover parameters to distribute the load over the network.

A Physical Resource Block (PRB) is the smallest unit of resources that can be allocated by a small cell eNB to a user in Long Term Evolution (LTE) {PRB}. One PRB occupies 180 kHz bandwidth in the frequency domain and 6 or 7 orthogonal frequency-division multiplexing (OFDM) symbols (i.e., one slot) in the time domain. The frequency part is composed of 12 consecutive sub-carriers of 15 kHz. Single sub-carrier with one symbol is defined as the resource element. Since LTE is capable of working with two types of duplex schemes frequency division duplexing (FDD) and time division duplexing (TDD), two different classes of radio frames are used. FDD uses Frame structure Type 1 (FST1), which is a widely used type; however; TDD uses Frame structure Type 2 (FST2). Each radio frame length is 10 ms and composed of 20 slots each is 0.5 ms long. Every two sequential slots compose one subframe, which is the time unit for scheduling users, denoted as the transmission time interval (TTI) (3GPP, 2016). Each cell has some available PRBs that is set according to the system bandwidth. For our system, cells adopt 20 MHz bandwidth, which corresponds to 100 PRBs.
The network UEs are classified into two classes. The values of the requested QoS metrics and frequency and duration of connections differentiate between the classes. UEs that belong to the same class have the same demand for minimum data rate, the maximum delay, etc. Each class j UE can report measurements of neighbouring small cells if the RSRP of that neighbouring is greater than a predefined threshold. Those are the candidate UEs for handover, and the reported small cells are the qualified candidate small cells. However, the different classes UEs select the neighbouring cells for handover based on several criteria. The small cells' eNBs enter some network usage characteristics in their database and make them available

to UEs. Examples of these characteristics could be the frequency and duration of the UEs' connections and (or) the QoS metrics (throughput, and delay).

### 4.2.2 System Model Constraints

This thesis introduces a dynamic UMLB for LTE-A Small Cell Networks with c-SON. We assume that c-SON has a powerful capability in terms of memory to handle a set of small cells' operations. However, there are still a few constraints that severely affect the network operation if not taken into account, which are as follows:

1. The proposed architecture considers a centralized controller that balances the load across the small cells; consequently, the network may collapse if the controller fails.
2. Handover delay includes a pre-handover time that the UMLB takes in searching and examining the target cells. Thus, pre-handover time must be in a range that keeps the handover delay less than the allowed values (3GPP, 2012a). Delays below 250 milliseconds (ms) are acceptable. Through simulation, we estimated the time needed for the proposed algorithm and found it to be 21.404 ms.

Previous work introduced in (Alexandris et al., 2016) discussed possible solutions for these challenges.

### 4.2.3 Cell Load Calculation

To accurately measure the load of a network cell, a proper method is adopted. Several ways have been used to represent the load of the cells, such as the number of users served by a cell and the load of the transport network. In this study, we adopt the resource block utilization ratio RBUR, which is the ratio of PRBs allocated by a small cell to that total available number of PRBs belong to that cell. The resource block utilization ratio directly limits the number of UEs that can be served by a specific data rate and delay constraints. The average RBUR of small cell s at time t over a time T is given by:

$$\overline{RBUR_{\ s}}(t) = \frac{\sum_{\tau\in(t-T,t)} \sum_{j\in J} I_{s,j}(\tau) * N_{s,j}(\tau)}{T * PRB} \tag{4.1}$$

where $I_{s,j}(\tau)$ is a binary indicator so that $I_{s,j}(\tau) = 1$ if user $j$ is served by small cells, $N_{s,j}(\tau)$ is the number of physical resource blocks assigned by small cell $s$ to UE $j$ at period $t$, and $PRB$ is the total number of resource blocks available at cell $s$. Note that each UE is served by only one small cell. Furthermore, all small cells have the same limited number of PRBs. Hence, the total allocated number of PRBs by a small cell $s$at time $t$ cannot exceed the maximum number of PRBs of the cell, $\sum_{j\in J} I_{s,j}(\tau) * N_{s,j}(\tau) \le PRB, \forall s$.

When the value of $RBUR$ reaches 1, the cells' resources are depleted and, any UE coming into this cell will either be disconnected or served by a lower data rate. In this research, we ignored any call admission control policy. Thus, when a new UE moves into an overloaded cell, it will be admitted by that cell, but the per-UE throughput in this cell will be affected. Hence, shifting UEs forcibly from the highly loaded cells to normal or lightly loaded cells is critical to mitigating the overload status.

### 4.2.4 Load Balancing Problem Formulation

Network performance determined by Key Performance Indicators (KPIs) indicates its QoS. Based on these KPIs, the c-SON identifies the optimum handover parameters for the edge-UEs and involving small cells to achieve a more stable network with the highest achievable QoS concerning load demand. Following is the KPI considered for the dynamic mobility load-balancing problem.

*Key Performance Indicator for Load Standard Deviation:*

In this work, a load standard deviation $\sigma$ is monitored, which determines the level of load balancing in the network at a time and is evaluated by using the load distribution in all small cells. The load standard deviation is defined as:

$$\sigma = \sqrt{\frac{\sum_{s \in S}(\overline{RBUR_{\ s}}(t) - RBUR_{Net}(t))^2}{S}} \tag{4.2}$$

Where $RBUR_{Net}(t)$ is the average $RBUR$ of a network of $S$ small cells at time $t$ during period $T$ as well. For simplicity, we omit the time symbols throughout the chapter.

The range of $\sigma$ is in the interval$[0, 1]$, with a lower value representing a highly balanced load distribution amongst all active small cells. Therefore, minimizing $\sigma$ is one of the objectives of this work to achieve a highly balanced load in the small cell networks.

## 4.3 Mobility Control Parameters

A network lets UEs report measurements of the signal quality of the serving and neighbouring cells either periodically or as event-driven reports. The signal qualities required to be measured by serving cells can be RSRP or RSRQ. However, LTE-A offers a set of event-driven measurement report mechanisms to minimize the signalling overhead in the network (Schmelz et al., 2011). Those events are performed by UEs for both the serving cell and the neighbouring cells. Therefore, we adopted some of them in our work for the network information-gathering phase and handover execution procedure.

### 4.3.1 LTE Events

LTE specified eight types of events a UE must report by Radio Resource Control (RRC) Connection Reconfiguration message. Events A1 to A6 are defined for intra-LTE mobility, while events B1 and B2 are tailored for inter-RAT mobility. Since we are dealing with intra-LTE mobility, we focus only on the A1 to A6 events.

If the criteria for a certain event have been satisfied for a predefined time, called time to trigger (TTT), the UEs perform intra-LTE event-triggered reporting. RSRP or RSRQ are the quantities used by the LTE network system for events measurement triggering and reporting. Some events are threshold-based events for which the network operator sets a predefined threshold.

For instance, the A1 event is triggered by UE when the serving cell becomes better than a given threshold, whereas the A2 event is triggered when the serving cell becomes worse than a predefined threshold. Event A4 is triggered by UE when a neighbouring cell becomes better than a predefined threshold. However, event A5 is triggered when the serving cell becomes worse than a given threshold, and a neighbouring cell becomes better than another given threshold.

On the contrary, events A3 and A6 are triggered using an offset value, which is a type of relative value regarding something such as the serving cell. Event A3 is triggered when a neighbouring cell becomes offset better than a primary serving cell, whereas event A6 is triggered when a neighbouring cell becomes offset better than a secondary serving cell. In our work, we will follow the standard in employing A3 event measurements for triggering and reporting handovers. Also, we adopt A4 event measurements to gather candidate edge-UEs and their corresponding neighbouring cells for the hastened handover.

### 4.3.2 A3 and A4 Events For load Shifting and Edge-UEs Finding

As stated above, event A3 is triggered based on the relative signal quality of the neighbouring cell. In other words, a UE triggers event A3 and reports the measurement to its serving cell when a neighbouring cell shows a better signal quality than the serving cell by some offset in dB. Thus, event A3 has been commonly used for triggering handovers in wireless networks. The small cell eNB configures the UE to measure the signal quality (RSRP) of the serving cell and neighbouring cells and trigger a handover when an "entry condition" has been maintained for a duration of time larger than $TTT$. The entry condition to trigger and report the A3 event measurements are expressed as follows:

$$M_n + O_{fn} + O_{cn} - Hyst > M_s + O_{fs} + O_{cs} + Off \tag{4.3}$$

where $M_n$ and $M_s$ are the RSRPs of the neighbouring cell and the serving cell, respectively. The $O_{fn}$ corresponds to the frequency-specific offset of the frequency of the neighbouring cell and the serving cell. However, $O_{fs}$ corresponds to the frequency-specific offset of the frequency of the serving cell. $O_{cn}$ is the cell-specific offset of the neighbouring cell, whereas $O_{cs}$ is the cell-specific offset of the serving cell. $Hyst$ is a hysteresis term for cell $s$, and $Off$ is the A3 event offset between the serving and neighbouring cells. Since we consider only intra-frequency handovers in this work, we ignore the inter-frequency parameters, i.e., $O_{fn}$ and $O_{fs}$. By adjusting the values of the parameters, $O_{cn}$, $O_{cs}$, and $Off$ of the above equation, it is possible to cause a particular UE currently served by cell $s$ to hand-over to the neighbouring cell $n$. Therefore, we can deliberately perform early, or late handovers based on the load status of the serving and neighbouring cells.

If we increase the value of $O_{cn}$ of a peculiar neighbouring or decrease the value of $O_{cs}$ of the serving cell, the cell range diminishes virtually, and hence the UEs will be handed over to the neighbouring cell. In contrast, reducing the value of $O_{cn}$ or increasing the value of $O_{cs}$ will increase the serving cell range which forces UEs to handover from the neighbouring to it; and hence the cell load increases. For example, in Figure 4.2 cell A is highly loaded with four UEs, and its neighbouring cells, B and C, both have less load with one UE and three UEs, respectively. At the overlap area of cells A, B, and C, there are two UEs, UE 3 and 4. Therefore, by increasing the $O_{cn}$ of cell B or cell C, UEs 3 and 4 handover from cell A to either cell B or cell C, and the network might become more balanced. As mentioned before, changing $O_{cs}$ of the serving cell might delay or speed the handover as well, but it may affect the cell range of all the neighbouring cells. Thus, the parameter $O_{cn}$ is more suitable since it shifts the load only to a neighbouring cell.

The c-SON can configure the handover parameters of event A3 to achieve a more balanced network. Still, the system needs information on the edge users' potential for early handovers. To that end, in our work, the A4 event is used to gather information on edge users of the overloaded cells. The UE triggers A4 event: when the RSRP of a neighbouring cell becomes better than a provided threshold:

$$M_n + O_{fn} + O_{cn} - Hyst > Threshold \quad (4.4)$$

Where $Threshold$ is the event A4's threshold. If entry conditions of the A4 event are satisfied by a UE, the UE can report measurements such as RSRP for the serving cell and neighbouring cell. The UE can report multiple neighbours, which means that such a UE has several$M_n s$,$O_{fn} s$ and$O_{cn} s$.

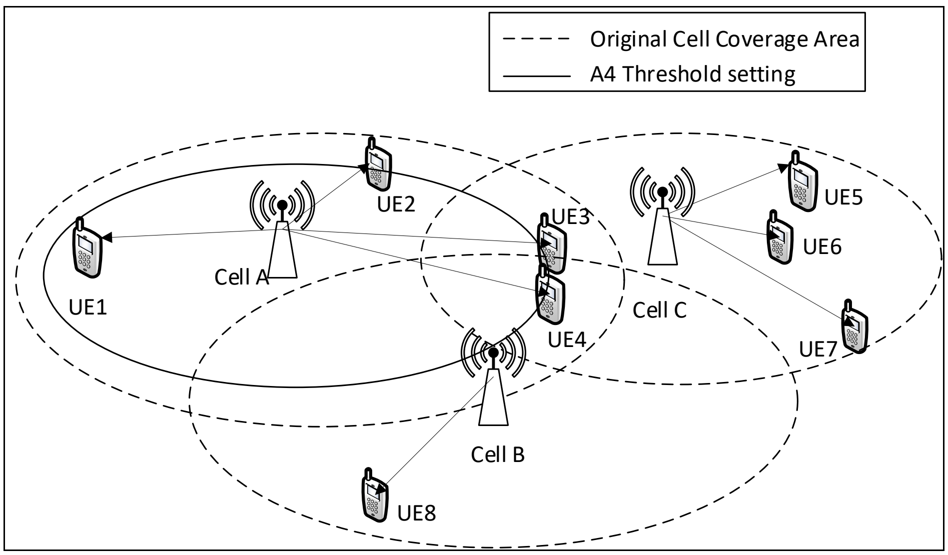



Figure 4.2 Original cell coverage and A4 event measurement reporting

The c-SON considers those neighbouring as neighbouring candidate cells. Small cells should set reasonable A4 event thresholds to collect edge-UEs' information and the candidate neighbouring small cells, both of which are required for load balancing. In Figure 4.2, let us assume that UEs 3 and four from cells C and B are outside of A4 event boundary of cell A and have reported RSRPs measurements as follows:

$$(M_B^{UE3}, M_B^{UE4}, M_C^{UE3}, M_C^{UE4}) + O_{fn} + O_{cn} - Hyst > Threshold_A \quad (4.5)$$

And in the meantime, the following conditions are true:

$$M_C^{UE3} > M_B^{UE3}$$
$$M_C^{UE4} > M_B^{UE4}$$

Based on measurement reports gathered from UEs three and four and load status of neighbouring cells $B$ and $C$, cell $A$ can hand UEs three and four over to cells $B$ and/or $C$ by increasing $O_{cn}$ for cell $B$ and $C$ when it becomes overloaded.

Thus, we utilize A4 event measurement reports for collecting information on edge-UEs that can report multiple neighbouring cells, and then c-SON selects the best neighbouring cell based on a combination of criteria. The c-SON collects edge-UEs' information $\varepsilon$ from small cells, which is $\varepsilon = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_S\}$, where $\varepsilon_S$ is the set of edge-UEs of small cell $s$.

## 4.4 Adaptive Utilization Threshold and Load Estimation

An adaptive threshold is needed to detect the overload state of the cells before starting the network balance by transferring suitable loads among cells.

### 4.4.1 An Adaptive Utilization Threshold for Load Status Detection

Two types of thresholds have been introduced to detect the status of the cells' load in the wireless networks: fixed and adaptive thresholds. The fixed threshold is not effective when applied to all scenarios as the load status of the mobile wireless network is dynamic and changes over space and time. Hence, an adaptive threshold has been introduced to adapt to the status of the network load. This method was proved to be better than the fixed threshold. Inspired by (Hasan et al., 2018), the adaptive threshold is defined to be the maximum of the average network load or a pre-defined fixed threshold:

$$TH_{Adp} = max(\overline{RBUR}_{NET}, TH_{INI}) \tag{4.6}$$

where $TH_{INI}$ is an initial fixed threshold at which the MLB algorithm triggers.

### 4.4.2 Calculation of User's required number of PRBs and After-Handover Load Estimation

We are considering only the downlink of an LTE system. We assume that the interference of eNB is the only interference in the network. For a pair of cells and UE $j$ in every time step, we assume that the Signal-to-Interference-plus-Noise ratio can be computed as:

$$SINR_{s,j} = \frac{P_s . L_{s,j}(d)}{N + \sum_{t \neq s} P_t . L_{t,j}(d) . RBUR_t} \tag{4.7}$$

where $P_s$ is the transmit power for a cell $s$, $L$ is the path loss mapping determined by the UE's location relative to a cell, $N$ is the thermal noise per PRB, and $RBUR_t$is the load of cell $t$. We assume that we can adopt the best modulation coding scheme (MCS) for a given SINR, which provides the highest data rate. That can be represented by the Shannon formula as follows:

$$r_{s,j} = B . log_2(1 + SINR_{s,j}) \tag{4.8}$$

Where $B$ is the bandwidth of one physical resource block (i.e., 180 kHz). Based on the data rate value demanded by an edge-UE, $R_{e_s^j}$, and the maximum achievable data rate using the assumed MCS at a given SINR $r_{s,j}$, the number of PRBs required by an edge-UE $e_s^j \in \varepsilon_S$ to keep up with the throughput requirement is estimated by:

$$N_{PRB}^{e_s^j} = \left\lceil \frac{R_{e_s^j}}{r_{s,e_s^j}} \right\rceil \tag{4.9}$$

Where $\lceil . \rceil$ is the ceil function. Note that the scheduler is the component that decides the number of PRBs allocated to UEs at a given time$t$. In our study, we are considering the Channel-Aware Quality of service scheduler (CQA). In addition to the SINR, we are taking into consideration the UEs' RSRPs measurements of the serving and neighbouring cells. The number of RBs required serving edge-UEs before and after handover differs because they experience different RSRP values from two different small cells. To perform a good load balancing, and before triggering the handover procedure, the algorithm should determine the current load of the edge-UEs to be handed over and estimate their load at the neighbouring cells.

Therefore, for a given number of PRBs $N_{PRB}^{e_s^j}$, the average load to serve the UE $e_s^j$ of small cell $s$ is denoted by $\rho\left(s, e_s^j\right)$ and calculated as following:

$$\rho\left(s, e_s^j\right) = \frac{N_{PRB}^{e_s^j}}{PRB_s} \tag{4.10}$$

Then, we can estimate the after-handover generated load by edge-UEs, $\hat{\rho}\left(k, e_s^j\right)$, at a neighbouring small cell $k$ by:

$$\rho\left(k, e_s^j\right) \approx \rho\left(s, e_s^j\right). \frac{Ms_{\left(s,e_s^j\right)}}{Mn_{\left(k,e_s^j\right)}} \tag{4.11}$$

where $Ms_{\left(s,e_s^j\right)}$ and $Ms_{\left(k,e_s^j\right)}$ are the RSRPs of serving cell $s$ and neighbouring cell $k$ measured by edge-UE $\mathrm{e}_\mathrm{s}^\mathrm{j}$ , respectively.

## 4.5 Proposed Work

In previous sections, we defined the system model, mobility parameter, and adaptive threshold and load estimation. Now, we will explain the proposed work. Two phases are performed: the data gathering phase and the UMLB algorithm phase.

### 4.5.1 Data gathering via networking monitoring

The c-SON supports automatic information gathering by monitoring the network. It periodically collects various information from the network small cells of the studied model. If the load of a small cell $s$ exceeds the computed adaptive threshold, the small cell is considered overloaded and should force a few UEs to handover to lightly load neighbouring cells.

Not only the cell load status is monitored, but also the information on the UEs that are at the edges of the small cells is gathered. For that, the c-SON tunes the A4-event threshold for each small cell and collects information on the UEs that are moving close to the edge. The information gathered is relative to the UEs' serving cells and their neighbouring cells. The A4-event threshold is computed and adaptively adjusted based on the RSRPs reported by A3-events that are performed by a set of UEs to cell $s$ for a predefined period,$T$. The RSRPs reported under A3 events by UEs in a serving cell $s$ are averaged, $\overline{Ms}$ for that serving cell. Then, the A4 threshold of serving cell $s$ is set to the average of $\overline{Ms}$ of the neighbouring cells is defined as:

$$TH_i = \frac{1}{||B_s||} \times \sum_{j \in B_s} \overline{Ms}_s \tag{4.12}$$

where $\overline{Ms}_s$ is the average RSRP reported by UEs at each serving small cell $s$; and $||B_s||$ is the neighbouring cell set of serving small cell $s$ that is reported by UEs based on A3 event measurement reporting during a time duration.

Next, the c-SON gathers measurement reports based on the A4-event threshold $TH_i$ from edge-UEs of each serving cell. These UEs are the candidate UEs to be shifted to the neighbouring lightly loaded cells when their serving cells become severely loaded. The c-SON creates a database for each cell based on the A4 events-triggered measurements. The database contains information on reporting UEs: their Identities, their serving and neighbouring cells' RSRPs, SINRs, etc. Let $\varepsilon_s = \{e_s^1, e_s^2, \dots e_s^{||\varepsilon_s||}\}$ indicates the set of edge-UEs that reported A4 measurements to serving cell$s$. They are listed in ascending order of the RSRP,$Ms$ of the serving cells. The c-SON creates another set that contains the neighbouring cells reported by edge-UEs of the serving cell. Let us assume that each edge-UE in set $\varepsilon_s$ can report multiple neighbouring cells $\mathcal{T}_{e_s^j} = \{\mathcal{T}^1, \mathcal{T}^2, \dots, \mathcal{T}^L\}$ where $L$ is the number of the candidate neighbouring cells for edge-UE$e_s^j$.

### 4.5.2 UMLB algorithm

The UMLB algorithm is run periodically by c-SON. To that end, the c-SON hand-overs candidate edge-UEs from the highly loaded serving cells to the normal or under-loaded neighbouring cells based on the utility function. First, all the small cells report their load information, $RBUR$ to the c-SON. Next, reporting small cells are sorted in descending order of$RBUR$.

Afterward, the algorithm compares the max load, $RBUR_{max}$ in the list with the predefined initial threshold. If the $RBUR_{max}$ is greater than the initial static threshold, the network is in overload status, and it requires an immediate load-balancing.

Then, for the algorithm to be adaptive to the network load status, we set the adaptive threshold, $TH_{Adapt}$, using equation 6. The current load of each small cell $\overline{RBUR}$ is compared with the adaptive threshold $TH_{Adapt}$ to detect the status of the load. If it is greater than the adaptive threshold, the cell is in an overload status, and accordingly, the c-SON algorithm performs load balancing.

The algorithm creates a new set $O$ that contains all overloaded cells such that $\overline{RBUR}_o \geq TH_{Adapt}$ for $o \in O$ and $O \subset S$ . Since the MLB algorithm mainly relies on the load status of the neighbouring cells and the UEs' positions of overloaded cells, the algorithm rearranges all overloaded cells in the set $O$ according to the remaining capacity of the neighbouring cells and the estimated edge-UEs' load at the neighbouring cells. To that end, we introduce the load balancing efficiency factor (LBEF) for an overloaded cell, which is defined as follows:

$$\psi_o = \sum_{k \subset B_o} min(\sum_{j} \rho(k,j), (1 - \widehat{RBUR}_k)) \tag{4.13}$$

where $B_o$ is the set of neighbouring cells reported by edge-UEs in cell $o$, $\rho(k,j)$ is the estimated after-handover edge-UEs load at the neighbouring cell, and $(1 - \widehat{RBUR}_k)$ indicates the number of RBs remaining at neighbouring cell $k$. Subsequently, the c-SON rearranges the set $O$ in decreasing order of $\psi_o$.

Then, the algorithm takes the overloaded cells one by one from the set and decreases its load to under-loaded neighbouring cells by handing-over the candidate edge-UEs. Each overloaded cell is computing the maximum load that can be moved to the target cells. That is to prevent the ultra-lightly loaded cells from becoming overloaded, and the serving cell from becoming underloaded. In other words, the UE is handed over if the maximum load moveable from the overloaded cell, $\tilde{\rho}_o$ is greater than the UE after-handover estimated load,$\rho(o, e_o^j)$. The moveable load is calculated by:

$$\tilde{\rho}_o = \overline{RBUR}_{NET} - \frac{1}{2}\rho(o, e_o^j) \tag{4.14}$$

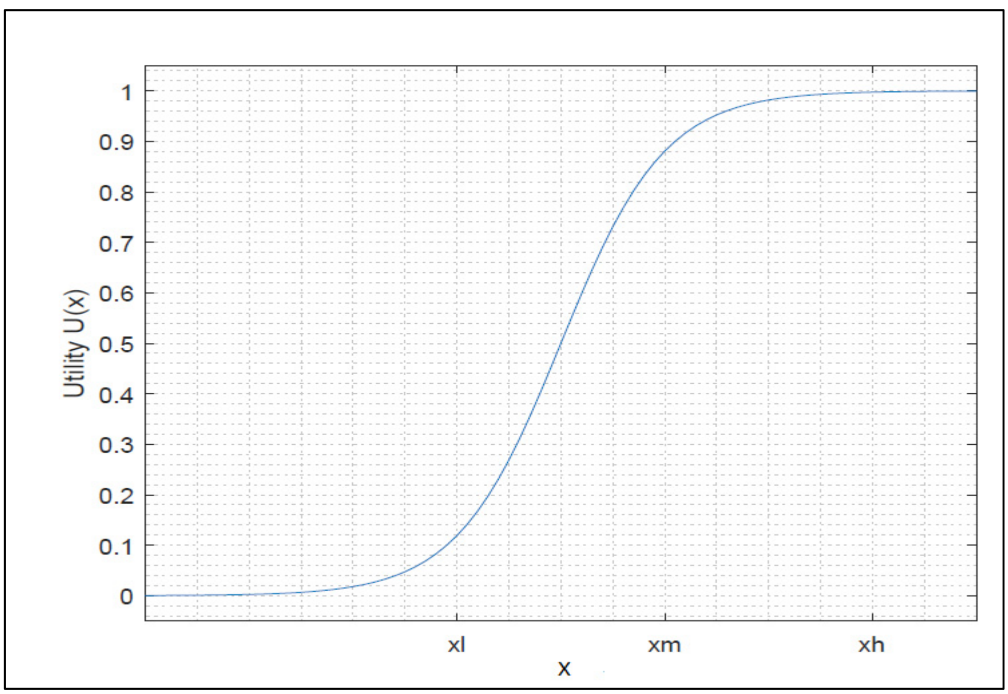
However, unlike the work that has been done in the literature, the handover process is based on the concepts of the utility function. In the following subsections, we model our system using the utility function. The algorithm of selecting a neighbouring small cell eNB for early handover initiates with calculating the user utilities.

We evaluate the UE utilities for each criterion for each neighbouring small cell candidate. The utility of a criterion is normalized to scale the interval $[0, 1]$, i.e. $u(x) \in [0,1]$, which indicates the UE satisfaction level from the criterion value offered by the small cell eNB. For instance, consider a scenario where a small cell eNB can serve a maximum of 18 Mbps and the requirement of a UE is in the range [5 Mbps, 30 Mbps]. With the help of the utility function, we can calculate the utility of the user for this data rate for this given eNB. In the second step, the c-SON evaluates the operator utility for the load criterion for each small cell eNB, which is a normalized function of a single criterion.

#### 4.5.2.1 UEs' utility calculation

We let UEs report multiple neighbouring cells based on event A4 measurements. In other words, let us look at the scenario of Fig.2 again, in which UEs 3 and 4 report small cells B and C if both $M_B^{UE3}$ and $M_B^{UE4}$ are greater than $Threshold_A$ of event A4. That means each UE report s two neighbouring cells to the c-SON.

We are considering a vector of $n$ criteria, $X = \{x_1, x_2, \dots, x_n\}$ with an associated vector of $n$ weights $W = \{w_1, w_2, \dots, w_n\}$ for the handover process. As stated above, the UEs concern about three criteria in this work: delay, data rate, and RSRP. The weights represent the UE's




preference level for a criterion $x_i$. The utility of a class $j$ UE for a small cell eNB $s$ and a

Figure 4.3 Utility calculation using sigmoid function predefined criterion $x_i$, denoted as $u_j^s(x_i)$, is calculated using a Sigmoidal (S-shaped) function, as presented in (Nguyen-Vuong et al., 2008). The utility is used to quantify the UE's satisfaction for a given criterion. Several utility function forms were examined whether they satisfy the required properties: twice differentiability, increasing function, concavity and convexity conditions. As a result of the examination, it was proved that only the sigmoidal functions can satisfy the required conditions of a utility function. The sigmoid function is shown in Figure 4.3 and expressed by the equation:

$$u(x) = \begin{cases} 0, & x_n < x_l \\ \dfrac{\left(\dfrac{x_n - x_l}{x_m - x_l}\right)^{\zeta}}{1 + \left(\dfrac{x_n - x_l}{x_m - x_l}\right)^{\zeta}}, & x_l \le x_n \le x_m \\ 1 - \dfrac{\left(\dfrac{x_h - x_n}{x_h - x_m}\right)^{\gamma}}{1 + \left(\dfrac{x_h - x_n}{x_h - x_m}\right)^{\gamma}}, & x_m \le x_n \le x_h \\ 1, & x_n > x_h \end{cases} \tag{4.15}$$

where $\zeta \geq max\left\{\frac{2(x_m - x_l)}{x_h - x_m}, 2\right\}$ and $\gamma = \frac{\zeta(x_h - x_m)}{x_m - x_l}$ are the parameters that determine the steepness of the utility curve, $x_n$ is the value obtained for the criterion $x$, $x_l$ is the minimum acceptable value for the criterion $x$, $x_h$ is the maximum desired value for the criterion $x$, and $x_m$ is a user-specific value that separates the satisfied from unsatisfied areas.

It is important to notice that equation (4.15) is defined for upward criteria for which the higher the values, the greater their utility (e.g., data rate, RSRP). However, we use $1 - u(x)$ for downward criteria for which the lower the metric value, the greater their utility (e.g., delay, load). Once the separate utilities from different criteria, the aggregate utility is computed as follows:

$$\mathfrak{U}^{\tau^l}_{e_o^j}(Z, W) = \prod_{i=1}^{n} \left[\mathfrak{u}^{\tau^l}_{e_o^j}(x_i)\right]^{w_i}, e_o^j \in J, \tau^l \in S, \tag{4.16}$$

$$\sum_{i} w_i = 1$$

In this form, the interactions/dependence among the considered criteria is considered. Also, it can take into consideration the UE preference weights for different criteria.

#### 4.5.2.2 Operator utility calculation

In the second step, for every candidate edge-UE, $e_o^j$ of an overloaded cell $o \in O$, we compute its load contribution in the current serving cell as $\rho(o, e_o^j)$ and estimate the after-handover load in each reported neighbouring cell as $\hat{\rho}(\tau^l, e_o^j)$. Hence, we can estimate the after-handover load of the neighbouring cell as:

$$\widehat{RBUR}_{\tau^l} = \overline{RBUR}_{\tau^l} + \hat{\rho}(\tau^l, e_o^j) \tag{4.17}$$

where $\overline{RBUR}_{\tau^l}$ is the estimated current load of the neighbouring cell.

After computing the current edge-UE load contribution and estimating the after-handover neighbouring cell load, we use the following formula to compute the operator utility for each neighbouring cell eNB and each class $j$ UE, which is expressed as follows:

$$u_{e_o^j}^{\tau^l} = \begin{cases} 1 - \widehat{RBUR}_{\tau^l}, & \overline{RBUR}_o - \rho(o, e_o^j) > \widehat{RBUR}_{\tau^l} < TH_{Adp}^{\tau^l} \\ 0, & Otherwise \end{cases} \tag{4.18}$$

Note that this equation is for the downward criterion for which the lower the metric value, the greater the utility. The load cost values are normalized to prevent it from dominating the handover utility function. Here, conditions of the operator utility restrict the release of the load from the overloaded cell so that this cell does not become underutilized and the neighbouring cell does not get overloaded either, and hence the algorithm does not enter into an infinite loop of load balancing. When the operator utility is zero, it means that the small cell is overloaded, and it must be eliminated in the elementary operator utility.

#### 4.5.2.3 Aggregate utility and handover

The overall utility for neighbouring small cell eNB $\tau^l$ and class $j$ UE is obtained by aggregating the UE and operator utilities for this small cell eNB. To that end, we apply the multiplicative aggregation form again to calculate the neighbouring cell eNB utility as follows:

$$\mathcal{U}_{e_o^j}^{\tau^l} = \mathfrak{U}_{e_o^j}^{\tau^l \, w_u} \times u_{e_o^j}^{\tau^l \, w_o}, \tau^l \in S, e_o^j \in J, \tag{4.19}$$

$$w_u + w_o = 1$$

where $w_o$ and $w_u$ represent the operator and user utility weights, respectively. Then, the best neighbouring cell eNB to handover the candidate edge-UE $\tau^l$ to is the one with the greatest utility value among all$\mathcal{U}_{e_o^j}^{\tau^l}, \tau^l \in S$.

Hence, the algorithm updates the related cell individual offsets as follows:

$$CIO = Mn_{(o, e_s^j)} - Ms_{(\tau^l, e_s^j)} + Hyst + \Delta \tag{4.20}$$

$$Ocn_{(o, \tau^l)} = CIO \tag{4.21}$$

$$Ocn_{(\tau^l, o)} = -CIO \tag{4.22}$$

Where $Mn_{(o, e_s^j)}$ is the RSRP of the overloaded serving cell $o$, $Ms_{(\tau^l, e_s^j)}$ is the RSRP of the target neighbour cell $\tau^l$ reported by $e_s^j$, $Hyst$ is a hysteresis parameter of cell $o$, $\Delta$ is the offset step-size, $Ocn_{(o,\tau^l)}$, and $Ocn_{(\tau^l,o)}$are the cell individual offsets for the target and serving cells, respectively. Notice that the offsets are always set symmetrical to prevent ping-pong. Eventually, the algorithm updates the serving and neighbouring cells' load information as follows:

$$\widehat{RBUR}_o = \widehat{RBUR}_o - \rho\left(o, e_o^j\right) \tag{4.23}$$

$$\widehat{RBUR}_{\tau^l} = \widehat{RBUR}_{\tau^l} + \hat{\rho}\left(\tau^l, e_o^j\right) \tag{4.24}$$

Characteristics such as the behaviour of the UE, metrics for QoS are dynamically updated so that the UEs read the updated values and recompute the utilities again. The whole above process applies to the overload cells list. The process for utility-based mobility load balancing is depicted in Algorithm 4.1.

Algorithm 4.1 Utility-Based MLB Handover Algorithm

```
Input: cellist, selected UE.
Output: a cell to receive the selected UE
MaxAggragate ← MIN
allocated_Cell ← None
TCells [] ← selected_UE.getNeighboringCells
foreach cell in TCell[] do
    Aggregate ←      get.cell.UEUtility  get.cell.OperUtility
If (Aggregate > MaxAggragate ) then
        MaxAggragate ← Aggregate
        allocated_cell ← cell
    endif
end for
  return allocated_cell
```

#### 4.5.2.4 Illustrative Scenario:

Consider the scenario shown in Figure 4.2. Let us suppose that UE3 is a class 1 UE being served by cell A and report cells B and C. By using values of class 1 users in Table 4.2 and equation (4.15); we define the terms of equation (4.15) as follows: delay: $x_l = 0$, $x_m = 0.5$, $x_h = 0.75$, the weight $= 0.22$, data rate: $x_l = 128$,$x_m = 256$, $x_h = 512$, the weight $= 0.38$, and RSRP: $x_l = -144$, $x_m = -100$, $x_h = -55$, the weight $= 0.4$.

Then, we define $x_n$ which might be the delay, data rate or RSRP values offered by cell B or cell C. For cell B: $x_n$ (Delay) = 0.4, $x_n$ (Data Rate) = 768, and $x_n$ (RSRP) $= -110$. For cell C: they are $x_n$ (Delay) = 0.3,$x_n$ (Data Rate) $= 200$, and $x_n$ (RSRP) $= -55$.

Now substitute them in the first part of the equation (4.15). We start with the delay for cell B. Since $x_l \leq x_n \leq x_m$ ($0 \leq 0.4 \leq 0.5$), we use the second part of the equation (4.15):

$$u_3^{\mathrm{B}}(Delay) = 1 - \frac{\left(\frac{x_n - x_l}{x_m - x_l}\right)^{\zeta}}{1 + \left(\frac{x_n - x_l}{x_m - x_l}\right)^{\zeta}}$$

We calculate $\zeta$ by $\zeta \geq max\left\{\frac{2(x_m - x_l)}{x_h - x_m}, 2\right\}$

$$\zeta \geq max\left\{\frac{2(0.5 - 0)}{0.75 - 0.5}, 2\right\} \geq max(4, 2) = 4$$

Thus,

$$u_3^{B}(Delay) = 1 - \frac{\left(\frac{0.4 - 0}{0.5 - 0}\right)^{4}}{1 + \left(\frac{0.4 - 0}{0.5 - 0}\right)^{4}} = 0.7094$$

Notice that we subtract it by one since delay criterion is a downward criterion.

Next, we do this again for data rate and because $x_n > x_h$, we use the last part of the equation (4.15) and hence:

$$u_3^{\mathrm{B}}(Data\ Rate) = 1$$

Next, we compute the UE3 utility for RSRP criterion. We note that$x_l \leq x_n \leq x_m$ $(-144 \leq -110 \leq -100$ ), and thus we use the second part of the equation (4.15):

$$u_3^{B}(RSRP) = \frac{\left(\frac{-110-(-144)}{-100-(-144)}\right)^{\zeta}}{1+\left(\frac{-110-(-144)}{-100-(-144)}\right)^{\zeta}}$$

And $\zeta$ is calculated by:

$$\zeta \geq max\left\{\frac{2(-100-(-144))}{-55-(-100)}, 2\right\} = \max(1.95,2) = 2$$

$$u_3^{\mathrm{B}}(RSRP) = \frac{\left(\frac{-110-(-144)}{-100-(-144)}\right)^{2}}{1+\left(\frac{-110-(-144)}{-100-(-144)}\right)^{2}} = 0.3738$$

Hence, UE3 utility from cell B is combined using equation (4.15) as follows:

$$U_j^{\mathrm{B}}\ (UE3\ utility\ comibation\ frorm\ cell\ B) = u_j^{\mathrm{B}}(Delay)^{w_{Delay}} * u_j^{\mathrm{B}}(DataRate)^{w_{Datarate}} * u_j^{\mathrm{B}}(RSRP)^{w_{RSRP}}$$

$$U_3^{\mathrm{B}} = 0.7094^{0.22} * 1^{0.38} * 0.3738^{0.4} = 0.6255$$

For cell C, we do the same steps as we did for cell B. For the sake of brevity, we brought the final answers as follows:

$$u_3^C(Delay) = 0.8852$$

$$u_3^C(Data\ Rate) = 0.24$$

Next, to calculate the RSRP UE3 utility, we use the third part of the equation (4.15):

$$u_3^C(RSRP) = 0.3970$$

Finally, we compute the UE3 combined utility from cell C as follows (equation (4.16)):

$$U_j^3\ (UE3\ utility\ comibation\ from\ cell\ C) = u_j^C(Delay)^{w_{Delay}} * u_j^C(DataRate)^{w_{Datarate}} * u_j^C(RSRP)^{w_{RSRP}}$$

$$U_3^C = 0.8852^{0.22} * 0.24^{0.38} * 0.3970^{0.4} = 0.3911$$

Now, let us calculate the operator utility. We assume that the cell A is 82% loaded, cell B current load is 70%, the current and estimated load for the UE3 at cell B is 10%, and the overload threshold is 81%. Note that we use the question marks intentionally to express specific conditions (e.g. 7 ?> 9 means is 7 value greater 9). Then, we check the condition of equation (4.18) as follows:

$$82 - 10\ >?\ 80\ + 10\ ?<\ 81$$

No, it's 72 < 90 > 81. Thus

$$u_3^B(Operator) = 0$$

We do the same to calculate cell C operator utility. We assume that cell A is 82% loaded, cell C current load is 45%, the current and estimated load for the UE3 at cell B is 10% and 20%, respectively, and the overload threshold is 81%. Then, we check the condition of equation (4.18) as follows:

$$82 - 10 \ >?\ 45 + 20 \ ?<\ 81$$

Yes, the condition is satisfied: 72 > 65 < 81

$$u_3^C(Operator) = \frac{1 - 65}{100} = 0.35$$

Finally, we compute the aggregate utility for UE3 using equation (4.16). $w_u,\ w_o$ are user and operator utility weights, respectively. Let's suppose$w_u \ = \ w_o = 0.5$, thus cell B utility is given by:

$$U_3^B(Aggregate\ utility) \ = 0.6255^{0.5} * \ 0^{0.5} = 0$$

And for cell C, the utility is given by:

$$U_3^C(Aggregate\ utility) \ = 0.3911^{0.5} * \ 0.35^{0.5} = 0.3699$$

We note that $U_3^C > U_3^B$, and hence UE3 is handed over to Cell C. We iterate these steps for all edge users and their reported cells.

## 4.6 Simulation

### 4.6.1 Simulation Environment

To study the proposed UMLB algorithm performance, we conducted a system-level simulation. A small cell network consisting of 10 small cells and 80 UEs is assumed in the simulation. UEs are split into two classes. Each small cell is assumed to use a bandwidth of 20 MHz. Hence, the number of total available resources is 100 PRBs. The transmission power is set to 24 dBm. The path loss was modelled as a non-line of sight propagation loss (Andersen et al., 1995). For allocating resources to the UEs, the channel and QoS aware (CQA) scheduler is used. We set the initial overload threshold to 0.75 for the proposed algorithm. The rest of the parameters are shown in Table 4.1. In the considered scenario, a full-buffer traffic model is used.

Regarding the initial UEs distribution over the network, 50% of UEs were static and non-uniformly distributed over the overlapping area of the small cells. For the sake of mobility, the remaining 50% of UEs were modelled with a circular way (CW) mobility model with a speed of 3.6 km/h and randomly distributed over the network coverage area.

Table 4.1 Simulation Parameters

| **Parameters** | **Values** |
|---|---|
| System bandwidth | 20 MHz |
| Transmission power | 24 dBm |
| Number of small cells | 10 |
| Inter-site distance (ISD) | 30m |
| Antenna mode | Isotropic |
| Number of UEs | 80 |
| Pathloss (NLOS) | $147.4 + 43.3\ log10(R)$ |
| Fading | Standard deviation 4 dB, log-normal |
| Scheduler | $CQA_{ff}$ |
| $CIO_{min}$ and $CIO_{max}$ | -6 dB, 6 dB |
| Hysteresis | 2 dB |
| $\Delta$ | 1 dB |
| Initial Threshold | 0.75 |

### 4.6.2 Calculation of user and operator utilities

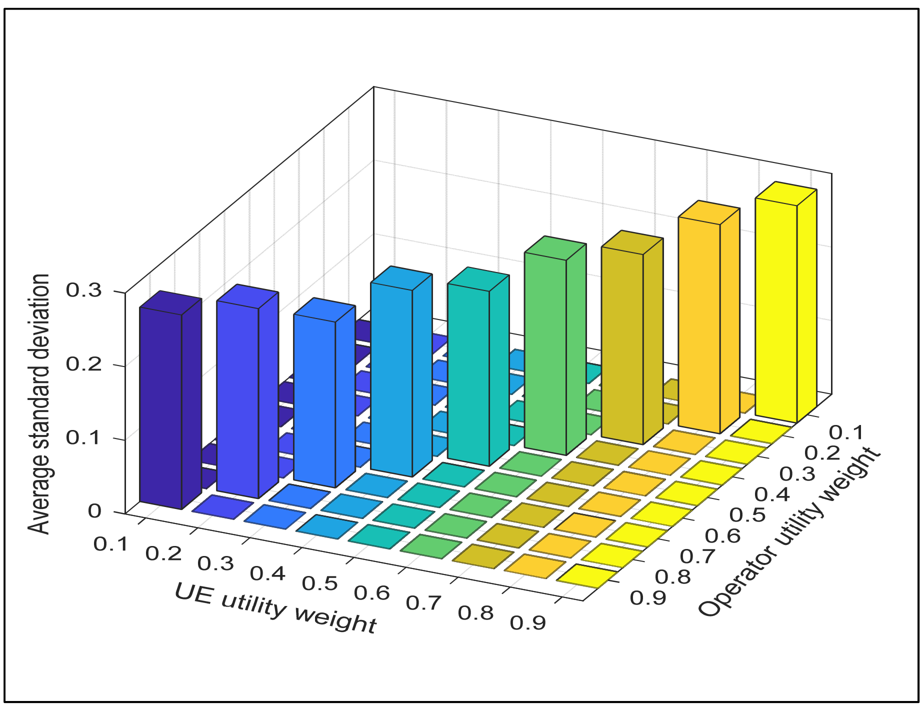
After gathering information from the cells' eNBs, detecting the overloaded cells and the candidate edge-UEs for handover, the UMLB algorithm calculates the UE utilities of the desired criteria. Table 4.2gives the minimum, mean, and maximum values requested by the user for each criterion, as well as their corresponding weights. Then, the utility of each criterion is evaluated using values in Table II and Equation (15). Besides, the eNBs' offered values for data rate and delay are calculated based on eNB statistics that are logged over a sliding end of an interval for the users connected to it. While the RSRP values are calculated based on the propagation model mentioned in the previous section.

The c-SON uses instant cell RBUR values and their maximum capacity to calculate the operator utilities using Equation (18).

Table 4.2 Criteria Values Requested by Users

| | **Requested values from user: min, mean, max/weight** | |
|---|---|---|
| | Class 1 user | Class 2 user |
| Delay (ms) | 0, 0.75, 0.9 / 0.22 | 0, 0.5, 0.75 / 0.35 |
| Data rate (kbps) | 128, 256, 512/ 0.38 | 256, 512, 1000 /0.33 |
| RSRP(dBm)×$10^{-10}$ | -144, -100, -55/ 0.4 | -144, -90, -44 / 0.32 |

The weights of the aggregation attributed to the operator and UE utilities influence the standard




deviation. We examined the potential combinations of the weights, as shown in Figure 4.4. It

Figure 4.4 Standard Deviation vs Operator and UE utility Weights is clear the standard deviation is at its minimum when the operator utility weight is 0.7, and the UE utility weight is 0.3, respectively.

### 4.6.3 Performance Evaluation Metrics

We investigated the performance of no-MLB, Fixed-MLB, W/O LBEF-UMLB, and LBEF-UMLB algorithms in terms of the standard deviation, the UE average data rate. The standard deviation is a metric used to measure the load distribution across the network. The effect of MLB algorithms on load distribution across the network was examined. Figure 4.5 shows the RBUR for the scenarios that do not implement MLB as well as for the ones with MLB algorithms. Cells are represented by coloured bars ordered from left to right. Figure 4.5 shows that cells 5 and 6 have a load greater than 0.9 for more than 50% of the operation time.

However, cells 4 and 7, which are neighbours of cells 5 and 6, respectively, have been underutilized at an RBUR less than 0.6 for 90% of the operation time. On the contrary, when the MLB algorithms are adopted, as shown in Figure 4.5 b, c and d, the highly loaded cells shifted some of their load to the lightly loaded cells. As a result, the load across small cells became more balanced. The figures show that RBUR of cells 4 and 7 became 0.9 for 82% and 58% of the time, respectively. Hence, for RBUR values greater than 0.9, the gap between the RBUR occurrence times for cells is reduced, which means the load became evenly distributed among the cells. The proposed UMLB algorithm reports a higher RB utilization since the proposed load balancing mechanism considers multiple targets when handing over UEs. Furthermore, we can notice that the LBEF-UMLB algorithm introduces a slight enhancement to the UMLB for RBUR values greater than 0.9. That is due to the capability of the algorithm in offloading the proper overloaded cells first. The cell with a lightly loaded neighbourhood has the priority to be offloaded. If this metric is ignored and the algorithm follows the classic sequence (starting with the maximum-loaded cell), some cells might not have the chance to shift some UEs, especially during the initial operation cycles.

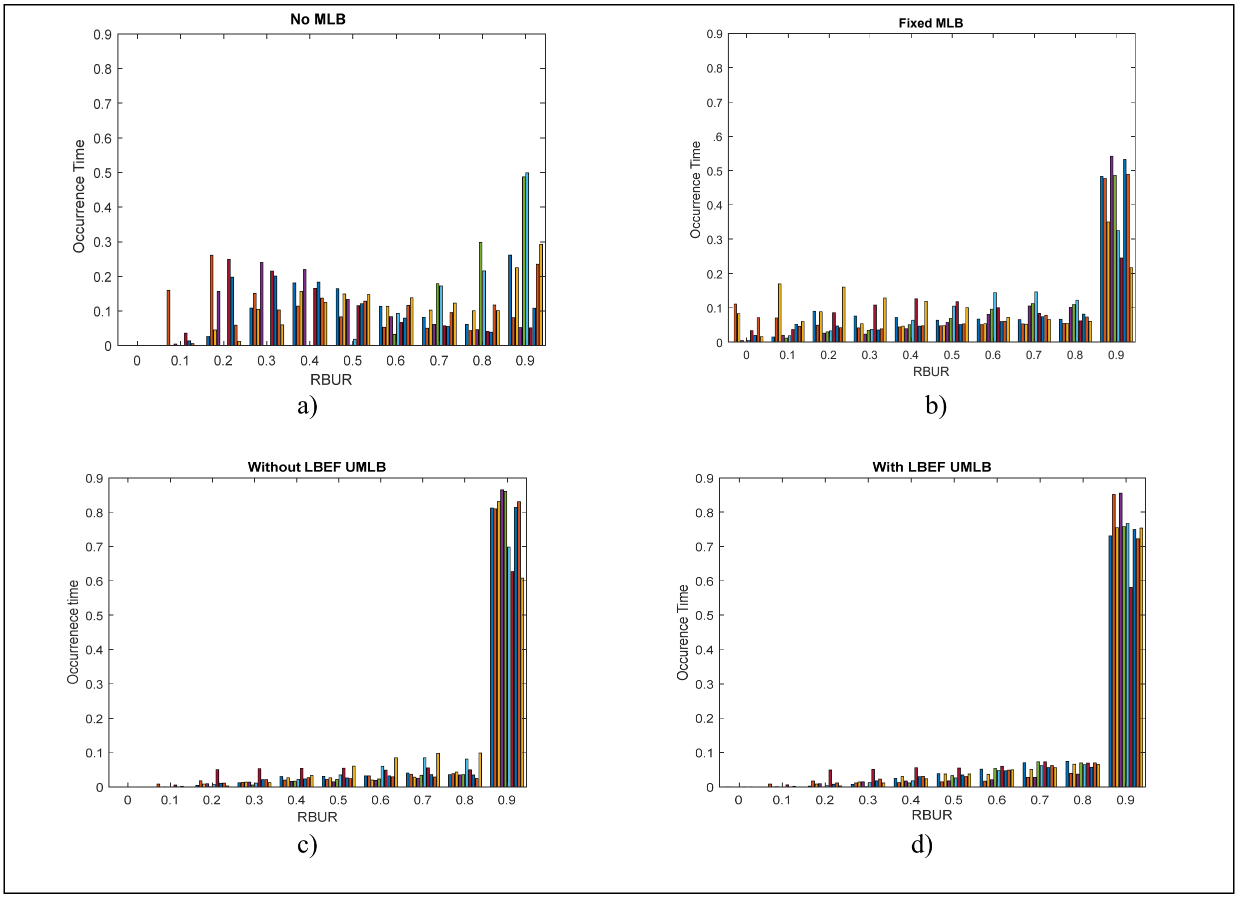



Figure 4.5 RBUR status of the network: (a) NO MLB algorithm (b) Fixed MLB (c) W/O LBEF UMLB (d) With LBEF UMLB

Figure 4.6 depicts the system performance for several MLB mechanisms in terms of load standard deviation across the small cell network.

The proposed algorithms achieved a smaller standard deviation compared with the other approaches. The proposed W/O LBEF-UMLB algorithm reduces the standard deviation by 75.86% and 74.07% for No-MLB and Fixed-MLB, respectively. The proposed LBEF-UMLB algorithm reduces the standard deviation by 77.58% and 75.92% for No-MLB and Fixed-MLB, respectively. Thus, the variance in load among small cells is lowered, and therefore, the system is more balanced.

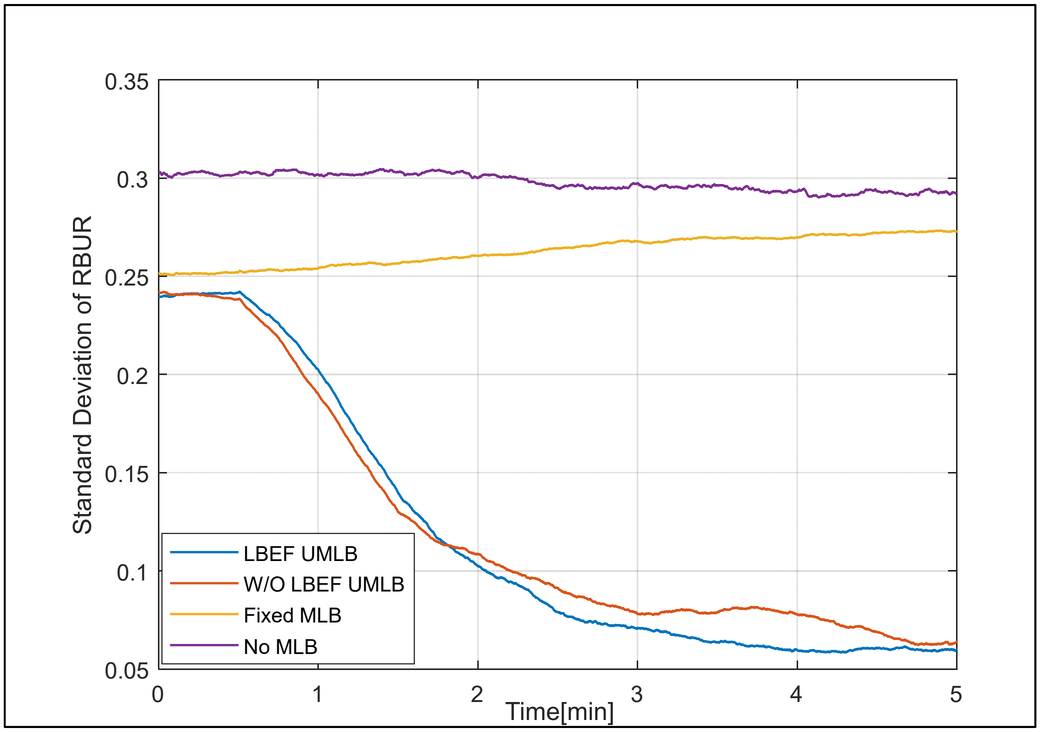



Figure 4.6 Standard deviation of RBUR among cells in the network

Moreover, we evaluated network performance in terms of the UE average rate. Figure 4.7 shows the average UE rate for multiple MLB approaches. Although there is always a trade-off between load balancing and throughput, the proposed algorithms increase the average UE data rate. That is because shifted UEs are allocated the required RBs at the neighbouring cell. If the UE in that overloaded cell is not handed over, it will experience a limited throughput due to a lack of RBs. The proposed LBEF-UMLB algorithms provide 40% of UEs with an average data rate of more than 1Mbps. On the other hand, approximately 1% of UEs can have an average data rate of 1Mbps when adopting No-MLB and Fixed-MLB algorithms.

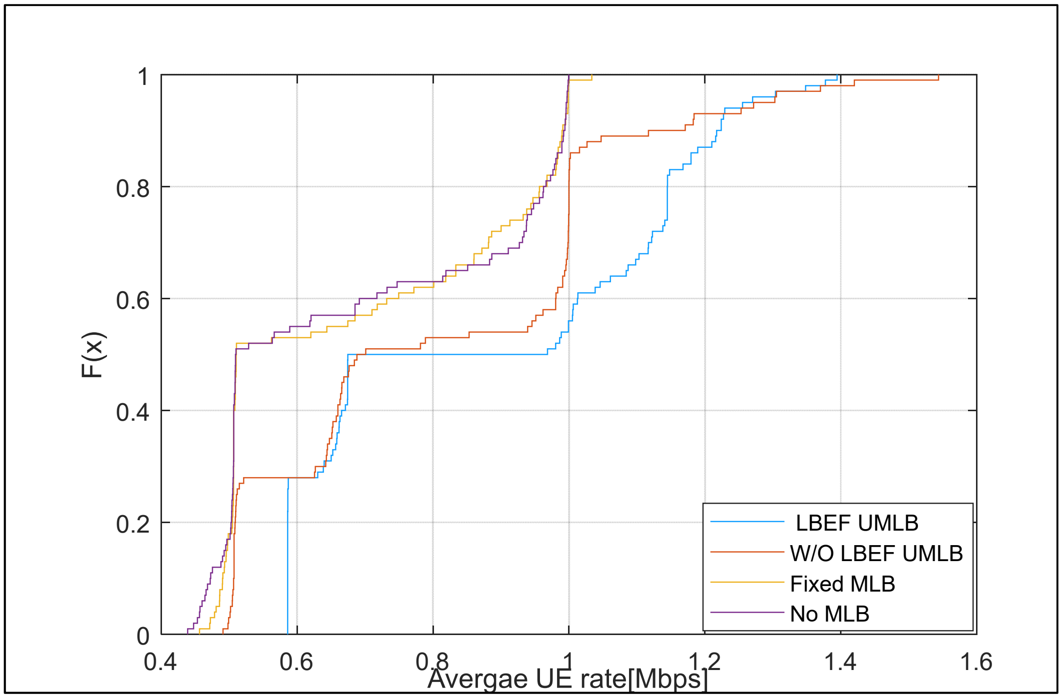



Figure 4.7 Average UE data rate [Mbps]

Since the proposed algorithms consider the UEs' preferences during the handover process, the UE average delay is enhanced compared to the No-MLB and Fixed-MLB, as shown in Figure 4.8. In this work, the delay is conceptualized as the difference between the achieved data rate and the required data rate.

Hence, it measures how much data is added to the transmission buffer. In other words, we measure delay as a fraction of the offered load; this allows us to measure how many slots the UE has experienced a delay as well as calculate the size of the UE's buffer queue. Hence, the less the cell is loaded, the less the delay is. As a result, we showed the impact of the cell load status on the UE's average delay. It is apparent from Figure 4.8 that the proposed algorithms exhibited a minimal delay in comparison to the other algorithms. The negative sign means that we are sending at a data rate higher than what is requested.

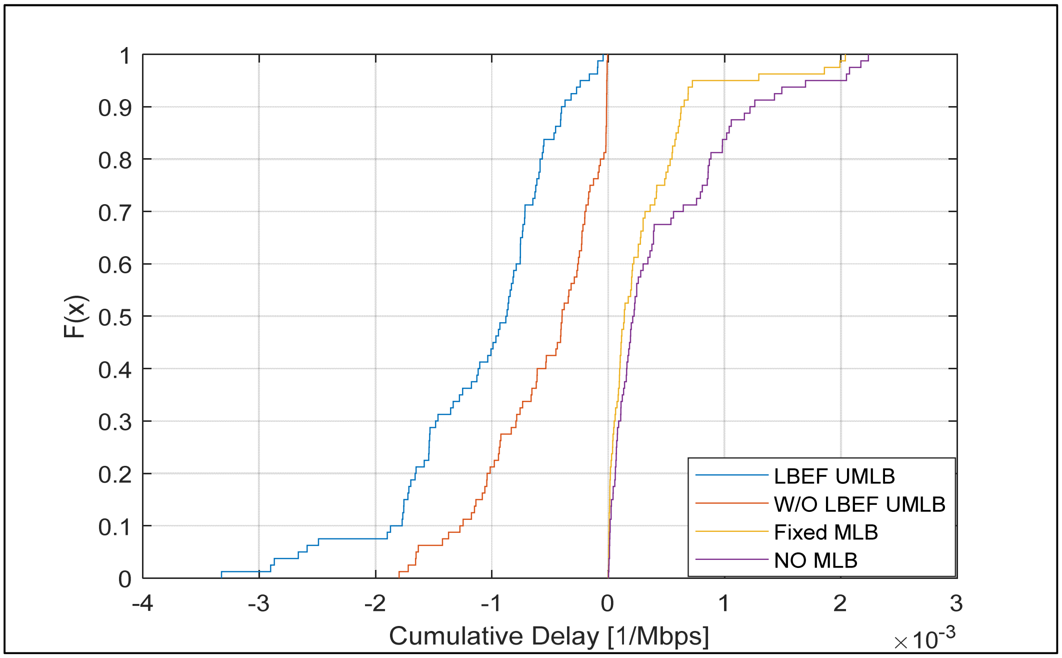



Figure 4.8 Average UE delay [1/Mbps]

## 4.7 Conclusion

To conclude all this, the load-imbalance across small cells in the network due to its low service area and mobility of UEs is examined. In this chapter, we introduced a UMLB algorithm and a new term named load balancing efficiency factor (LBEF). The UMLB balances the load across a small cell network by considering the operator utility and the user utility for the handover process. The operator utility is calculated for each potential handover based on the load of the neighbouring small cells and edge UEs of those neighbouring cells. Whereas, the user utility calculation is based on the sigmoid function by considering different criteria (i.e., RSRP, data raet and delay). Also, the LBEF considers a load of neighbouring cells and the edge-user equipment for each overloaded cell. This factor specifies the sequence of overloaded cells for the UMLB algorithm operation. The simulation results show that the UMLB minimizes standard deviation with a higher average-UE data rate when compared to existing load balancing algorithms. Therefore, a well-balanced network is achieved. However, implementing such UMLB algorithm leads to a greater number of performed handovers. This is considered as one of the costs of an MLB traditional algorithms. The work in the next chapter is to study the number of handovers required for the MLB process and minimize them while attaining the minimum standard deviation over the cells' loads.

# CHAPTER 5

# MOBILITY LOAD BALANCING WITH HANDOVER MINIMIZATION

## 5.1 Introduction

Mobility of UEs in a small cell network with low service area cells may cause load-imbalance across the cells in the network. The shortage of resources in the overloaded small cells leads to poor QoS when UEs intend to enter those cells though they have lightly loaded neighbouring cells. Consequently, resources of the unloaded cells remain underutilized though some overloaded neighbouring cells cannot meet the QoS requirements.

MLB distributes the load among small cells by adjusting the mobility parameters forcibly (i.e., handover parameters) according to their load statuses. To shift the candidate UEs, the cell individual offsets (CIO) of the serving and neighbouring cells are adjusted by UEs based on the reported measurements. Consequently, a greater number of handovers is required. This is one of the costs for the MLB operation that may affect the network performance severely (Oh et al., 2016). Thus, the network needs proper load balancing mechanisms to satisfy the QoS requirements with the minimum number of handovers.

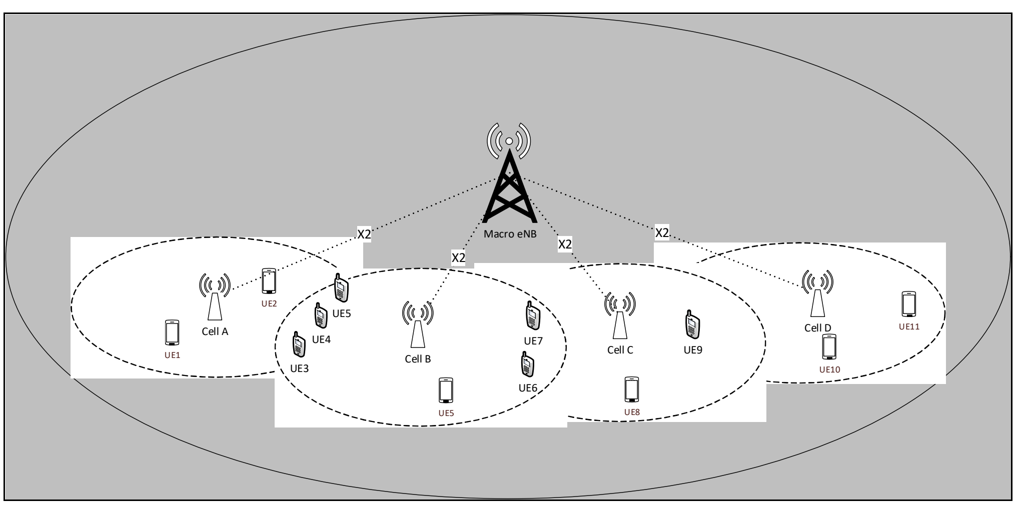



Figure 5.1 Heterogeneous Radio Access Network Architecture

In this chapter, we introduced a Utility-based Mobility Load Balancing algorithm with Handover minimization (UMLB-HO) by considering not only the edge UEs but also the non-edge UEs for overloaded small cells during the load balancing process. The main objective for any overloaded small cell during the MLB is to determine the fast-moving non-edge UE and transfer them to the underloaded Macro cell. Moreover, the small cell will determine the fast-, slow or very slow-moving edge-UEs during the MLB and transfer them to either under loaded neighbour small cell or Macro cell.

We define four important terms to specify whether the UE is fast-, slow or very slow-moving in order to determine the best handover decisions for each UE. First, the remaining service time (RT) is the estimated time a UE is going to stay in the serving cell. Second, the service Time at Target Neighbor (STN) that is the estimated time a user is going to stay at the neighbour cell. Last, the UE's session time (ST) is the needed time by the UE to use a certain application, while the remaining session time (Vesterstrom et al.) is the time equivalent to the remaining load that, a UE would impose in the near future.

Therefore, the UE is a non-edge fast moving UE if his RST is greater than its RT, thus he is going to be handed over to the Macro cell instead of small cells to avoid the unnecessary frequent handovers among the small cells. However, the UE (edge) is considered as a fast-moving if his session time is greater than the RT plus the STN, thus the UE will be handed over to the macro cell to avoid unnecessary frequent handovers. Moreover, the UE (edge) is a slow-moving UE when his RST is greater than the RT and less than the RT plus the STN, therefore; he will be handed over to a neighbour small cell. While the UE (edge) is considered as a very slow-moving UE if his RST is less than his RT, hence he will not be handed-over because his session will end before he leaves the serving cell, thus unnecessary handover is avoided.

For example, as shown in Figure 5.1, small cell A is overloaded with a load of 80%, whereas small cells B, C, and D have load values of 55%, 65%, and 60%, respectively. The adaptive overload threshold is assumed at 65%. For small cell A, let us assume that UE1 and UE5 are fast, UE2 is static, UE3 is very slow, and UE4 is slow. If the UMLB algorithm is adopted, for cell A, candidate UEs are listed as follows {UE3, UE5, UE4} based on RSRP values. For simplicity, the average load consumed by each UE is assumed 5%. Therefore, UE3 and UE5 are transferred to cell B, while UE4 is not transferred because cell B is going to be overloaded. Thus, the load of cells A and B become 70% and 65%, respectively. On the other side, if the UMLB-HO algorithm is adopted, for cell A, candidate UEs are listed as follows {UE5, UE1, UE4}. The algorithm will start with UE3 that will be not transferred to cell B because it is considered as very slow-moving and thus unnecessary handover is avoided. Then, UE5 is handed over to the Macro cell because it is considered as fast-moving and thus frequent handovers among the small cells are avoided (from cell A through cell D). Next, UE1 is handed over to the Macro cell because it is considered as a fast-moving UE. Finally, UE4 is transferred to cell B because it is a slow-moving UE. Therefore, the load of cells A and B becomes 65%, and the number of handovers performed is three using UMLB-HO instead of seven (frequent handovers from UE1, and UE5, and UE3 handover) with UMLB. Consequently, the network has a less standard deviation with a minimum number of handovers.

In general, the UMLB-HO algorithm is just an extension of the UMLB algorithm that considers the operator utility and the user utility explained in the previous chapter. However, the UMLB-HO aims to balance the load among the network with the minimum number of required handovers.

## 5.2 System Model

This section defines the network model that will be investigated throughout this chapter. Next, the most important system parameters and constraints are determined.

### 5.2.1 Network Model

First, we introduce the network model, its parameters and assumptions. In this subsection, we investigate a heterogeneous network of a macro cell $N$ and several small cells indicated by the set $\mathcal{S} = \{1,2 \ldots S\}$, as depicted in Figure 5.1. It also consists of a set of $\mathcal{U} = \{1,2, \ldots, U\}$ users. The entire network belongs to the same operator and small cells operate in an open access mode. We consider that for having core network functionalities, the cells are connected to the Evolved Packet Core (EPC), more specifically to the Mobility Management Entity (MME) and the Serving Gateway (S-GW) using $S1$ interface. Furthermore, the cells are interconnected via $X2$ interface, which enables the cells to directly communicate with each other and perform functionalities such as handover (Lobinger et al., 2011). Therefore, the users in the network can switch among the cells. For small cells, the centralized SON (c-SON) is adopted. In c-SON, some optimization functions are executed at the Operation and Management system (OAM), while others are executed at eNBs.

The c-SON periodically gathers information form small cells and uses them, if any overloaded cell is detected, to optimize and update the small cells' handover parameters to distribute the load over the network with the minimum number of handovers.

### 5.2.2 Load Balancing Problem Formulation

Network performance determined by Key Performance Indicators (KPIs) that indicates its QoS. Based on these KPIs, the c-SON identifies the optimum handover decisions for the UEs and involving small cells to achieve a more stable network with the highest achievable QoS concerning load demand.

The objective of this chapter is to balance the load across the small cells and maximize the QoS of the network. To that end, our problem is to minimize the load variance among the cells with the minimum number of handovers. Thus, the problem can be represented as:

$$MAX\ (QoS(t)) = MIN\ (\sigma(t), HO(t)) \tag{5.1}$$

$$Subject\ to\ C1\colon \sum_{j \in J} I_{s,j}(t) * N_{s,j}(t)\ \leq PRB,\ \forall s \tag{5.2}$$

$$C2: \sum_{j \in J} I_{s,j}(t) = 1, \forall u \tag{5.3}$$

where $\sigma(t)$ is the load standard deviation, and $HO(t)$ is the number of handovers required to balance the network load. $I_{s,j}(t)$ is a binary indicator so that $I_{s,j}(t) = 1$ if user $j$ is served by small cell $s$, $N_{s,j}(t)$ is the number of physical resource blocks assigned by small cell $s$ to UE $j$ at period $t$, and $PRB$ is the total number of resource blocks available at cell $s$. Therefore, the constraint $C1$ represents that the total allocated number of PRBs by a small cell $s$ at time $t$ cannot exceed the maximum number of PRBs of the cell. And constraint $C2$ states that the UEs are not allowed to be associated with multiple cells at one time.

## 5.3 Session Time Estimation and Remaining Service Time:

### 5.3.1 The User's Remaining Session Time (Vesterstrom & Thomsen)

The session duration (ST) for a UE depends on the application size, minimum and maximum requested data rates, and the data rate offered by the eNB. Thus, it is calculated as follows:

$$ST = \frac{L}{r} \tag{5.4}$$

Where L is the data size of the application in bits, and r is the offered data rate in b/s. Research in (Albasheir, 2016) presented detailed information about the Session Time and the factors that affect it. While the remaining session time (Vesterstrom & Thomsen) is the time equivalent to the remaining load that a UE would impose in the near future. Therefore, the RST can be represented as:

$$RST = ST - CT \tag{5.5}$$

Where consumed time (CT) is time equivalent to consumed load that a UE imposed in the past. Note that the eNB has the capability to evaluate the $CT$ at any time during the UE session time.

### 5.3.2 Remaining Service Time (RT)

The remaining service time is estimated for edge- and non-edge UEs using two similar methods. For a non-edge UE, at first, the system estimates the cell range in terms of RSRP of A3 events, $R_D$. The RSRPs of serving small cells ($M_{A3_S}$) are gathered from the A3 event measurement reports. Hence, the cell boundary can be calculated by averaging the total $M_{A3}$s for $k$ measurement reports as follows:

$$R_D = \frac{1}{k}\sum_{i=1}^{k} M_{A3_S} \tag{5.6}$$

When the received RSRP at a UE starts decreasing, the remaining service time is estimated. Thus, for a UE, any decrease in the received RSRP from the serving cell is observed.

UEs can report the received RSRPs from the serving cell based on the A2 event when the signal quality of the serving cell becomes worse than a predefined threshold. Therefore, the A2 event can be used to monitor a change in RSRP for non-edge UEs.

The remaining service time of a UE is estimated by monitoring the event A2 based RSRP measurements. We consult tow measurement reports under the event A2. When the UE moves outward, it reports that $R_{p_1} > R_{p_2}$, where $R_{p_1}$ and $R_{p_2}$ denote the RSRP measurement reports at times $t_1$ and $t_2$, respectively. Hence, the RSRP reduction rate, $\rho$, can be calculated over the interval report, $\Delta_t = t_1 + t_2$, as following:

$$\rho = \frac{1}{\Delta_t}(R_{p_2} - R_{p_1}) \tag{5.7}$$

Thus, the remaining service time of a non-edge UE can be calculated by:

$$\tau_{non-edgeUE} = \frac{1}{\rho} . (R_{p_2} - R_D) \tag{5.8}$$

For an edge UE, we adopt the A4 event as the first threshold ($R_{A4}$) that is corresponding to the first RSRP A2 event measurement report $R_{p_1}$. In addition, a new measurement report, with RSRP of $R_{A_3A_4}$, is added to be equivalent to the second RSRP A2 event measurement report $R_{p_2}$. Thus, we can calculate the reduction rate$\rho$. Therefore, the remaining service time of a particular edge UE can be estimated as:

$$\tau_{edgeUE} = \frac{1}{\rho} . \left(R_{A_3A_4} - R_D\right) \tag{5.9}$$

### 5.3.3 Service Time at Target Neighbor (STN)

The STN is the expected time that a UE is going to spend at the target neighbouring cell. This time is calculated based on the average speed of the UEs (V) (according to the mobility model adopted) and the coverage range of the small cell (D) as follows:

$$STN = D/V \tag{5.10}$$

## 5.4 Proposed Work

In previous sections, we defined the system model, session time estimation, and remaining service time. Now, we will explain the proposed work. Two phases are performed: the data gathering phase and the UMLB-HO algorithm phase.

### 5.4.1 Data gathering via network monitoring

The c-SON monitors the network and periodically collects various information from the network. If the load of a small cell $s$ exceeds the computed adaptive threshold, the small cell is considered overloaded and should force a few UEs to handover to lightly load Macro or neighbouring cells.
Next, the c-SON gathers measurement reports from the UEs of each serving cell. These UEs are the candidate UEs to be shifted to the lightly loaded cells when their serving cells become severely loaded. Based on these measurements, the UEs are classified into three sets: fast, slow and very slow UEs.

Let $u_{E-fast} = \{u^1_{E-fast}, u^2_{E-fast}, \dots\}$, $u_{NE-fast} = \{u^1_{NE-fast}, u^2_{NE-fast}, \dots\}$ indicate the set of fast-moving Edge- and Non-Edge UEs, respectively, and set $u_{fast} = \{u_{E-fast}, u_{NE-fast}\}$. Let us set $u_{slow} = \{u^1_{slow}, u^2_{slow}, \dots\}$ indicates the set of slow-moving UEs and set $u_{very-slow} = \{u^1_{very-slow}, u^2_{very-slow}, \dots\}$ indicates the set of very slow-moving UEs. Since the fast and slow UEs are the candidate UEs to be transferred, we create a new set $u_{candidateUEs} = \{u_{fast}, u_{slow}\}$. They are listed in ascending order of the RSRP, $Ms$, of the serving cell $s$. The c-SON creates another set that contains the neighbouring cells reported by edge-UEs of the serving cell. Let us assume that each edge-UE in the UE sets can report multiple neighbouring cells $\mathcal{T}_{u_{slow}} = \{\mathcal{T}^1, \mathcal{T}^2, \dots, \mathcal{T}^G\}$ where $G$ is the number of the candidate neighbouring cells for edge-UE $e^j_s$.

### 5.4.2 UMLB-HO algorithm

The UMLB-HO algorithm is run periodically by c-SON. To that end, the c-SON hand-overs candidate UEs from the highly loaded serving cells to the normal or under-loaded Macro or neighbouring small cells based on the utility function and the UEs' remaining service and session times. First, all the small cells report their load information, $RBUR$ to the c-SON. Next, reporting small cells are sorted in descending order of $RBUR$.

Afterward, the algorithm compares the max load, $RBUR_{max}$ in the list with the predefined initial threshold. If the $RBUR_{max}$ is greater than the initial static threshold, the network is in overload status, and it requires an immediate load-balancing.

Then, for the algorithm to be adaptive to the network load status, we set the adaptive threshold, $TH_{Adapt}$, using equatin (4.13). The current load of each small cell $\overline{RBUR}$ is compared with the adaptive threshold $TH_{Adapt}$ to detect the status of the load. If it is greater than the adaptive threshold, the cell is in an overload status, and accordingly, the c-SON algorithm performs load balancing.

The algorithm creates a new set $O$ that contains all overloaded cells such that $\overline{RBUR}_o \geq TH_{Adapt}$ for $o \in O$ and $O \subset S$. Subsequently, the c-SON rearranges the set $O$ in decreasing order of the efficiency factor calculated by equation (4.13).

Then, the algorithm takes the overloaded cells one by one from the set and decreases its load to the Macro or under-loaded neighbouring small cells by handing-over the candidate UEs. We classify the UEs (fast, slow or very slow) in every overloaded cell. The non-edge UE is a fast-moving UE if he satisfies the following condition:

$$RST > RT \tag{5.11}$$

However, the edge UE is considered as a fast-moving if the following condition is satisfied:

$$RST > (RT + STN) \tag{5.12}$$

The edge UE is considered as a slow-moving UE if the following condition is satisfied:

$$RT < RST < (RT + STN) \tag{5.13}$$

Finally, the edge UE is a very slow-moving UE when the following condition is satisfied:

$$RST < RT \tag{5.14}$$

Then, the algorithm takes the overloaded cells one by one from the set and decreases its load to under-loaded Macro or neighbouring small cells. Each overloaded cell is computing the maximum load that can be moved to the target cells using equation (4.14).
First, the algorithm starts with the very-slow edge UE set. But no hand-over will be performed since their session will end before they leave the serving cell, thus unnecessary handover is avoided.

Next, the algorithm defines the load of the Macro cell as $RBUR_{Macro}$, and assumes that the Macro cell has a fixed threshold, which is set by the operator, for its load status as $Th_{Macro}$. Then, it starts handing fast UEs from $u_{candidateUEs}$ set to the Macro cell according to the following equation:

$$RBUR_{Macro} + \hat{\rho}_{u_{fast}} < \ Th_{Macro} \tag{5.15}$$

Where $\hat{\rho}_{u_{fast}}$ is the estimated load of the $u_{fast}$ at the Macro cell. Consequently, frequent handovers across small cells will be avoided.

Then, it transfers the slow UEs from $u_{candidateUEs}$ set to a neighbour underloaded cell according to the UMLB algorithm introduced in the previous chapter.

Finally, the algorithm updates the related cell parameters (Cell loads, mobility parameters, etc.). The process for UMLB-HO is depicted in Algorithm 5.1.

Algorithm 5.1 Utility-Based MLB Handover Algorithm

Get RBUR information from macro cell and small cells;
Detect overloaded small cells and create set O;
Sort O according to efficiency factor;
Get and classify UEs according to equations (5.11), (5.12), (5.13), and (5.14)
**for** all $o \in O$ **do**
Calculate moveable load according to equation (5.14)
    **for** all $u_{veryslow}$ **do**
        No handover to any cell.
    **end for**
    **if** $RBUR_{Macro} < Th_{Macro}$ **then**
        **for** all $u_{fast}$ **do**
            **if** $RBUR_{Macro} + \hat{\rho}_{u_{fast}} < Th_{Macro}$ **then**
                Handover UE to Macro cell
            **end if**
        **end for**
    **end if**
    **for** all $u_{slow}$ **do**
        Execute UMLB algorithm
    **end for**
**end for**

## 5.5 Simulation

### 5.5.1 Simulation Environment

To study the proposed UMLB-HO algorithm performance, we conducted a system-level simulation. A heterogeneous network consisting of 1 Macro cell and 10 small cells with 80 UEs is deployed. UEs are split into two classes. The first class UEs are using a Video Streaming 240p application with data size 5.8 MB with requested data rate 1000,512,256 kbps that represent maximum, mean and minimum values, respectively. Whereas the second class UEs are using an Internet Browsing application with data size 2 MB with requested data rate 512,256,128 kbps that represent maximum, mean and minimum values, respectively. Each cell is assumed to use a bandwidth of 20 MHz. Hence, the number of total available resources is 100 PRBs. The transmission power is set to 46 dBm for the Macro cell and to 24 dBm for the small cell. The path loss was modelled as a non-line of sight propagation loss (Andersen et al., 1995). For allocating resources to the UEs, the channel and QoS aware (CQA) scheduler is used.

Table 5.1 Simulation Parameters

| **Parameters** | **Values** |
|---|---|
| Macro Tx | 46 dBm |
| Small Cell Tx | 24 dBm |
| System bandwidth | 20 MHz |
| Number of small cells | 10 |
| Inter-site distance (ISD) | 30m |
| Number of UEs in small cells | 80 |
| Macro Background Load | 0.2 |
| Pathloss (N-LOS) | $147.4 + 43.3\,log10(R)$ |
| Fading | Standard deviation 4 dB, log-normal |
| Scheduler | $CQA_{ff}$ |
| $CIO_{min}$ and $CIO_{max}$ | -6 dB, 6 dB |
| Hysteresis | 2 dB |
| $\Delta$ | 1 dB |
| Small cell Initial Threshold | 0.75 |
| Macro cell Load Threshold | 0.8 |
| Video Streaming 240p Data Size | 5.8 MB |
| Internet Browsing Data Size | 2 MB |
| Video Streaming 240p Requested Data rate (kbps) | 256, 512, 1000 |
| Internet Browsing Requested Data rate (kbps) | 128, 256, 512 |

We set the initial overload threshold for the small cells to 0.75, while we set a fixed overload threshold for the Macro cell to 0.7. The rest of the parameters are shown in Table 5.1. Here, we assume that the fixed data size traffic arrives at a fixed time interval per UE – in the downlink. Moreover, delay times (queuing, transmission, etc.) are neglected for simplicity.

UEs were modelled with a random walk mobility model with a speed of 3.6 km/h and randomly distributed over the network coverage area of small cells.

### 5.5.2 Performance Evaluation Metrics

Figure 5.2 depicts the system performance for several MLB mechanisms in terms of load standard deviation across the network. The proposed algorithm achieved a smaller standard deviation compared with the other approaches. The proposed UMLB-HO algorithm reduces the standard deviation by 87.54%, 86.25% and 38.21% for No-MLB, Fixed-MLB, and WUMLB, respectively. Thus, the variance in load among small cells is lowered, and therefore, the system is more balanced.

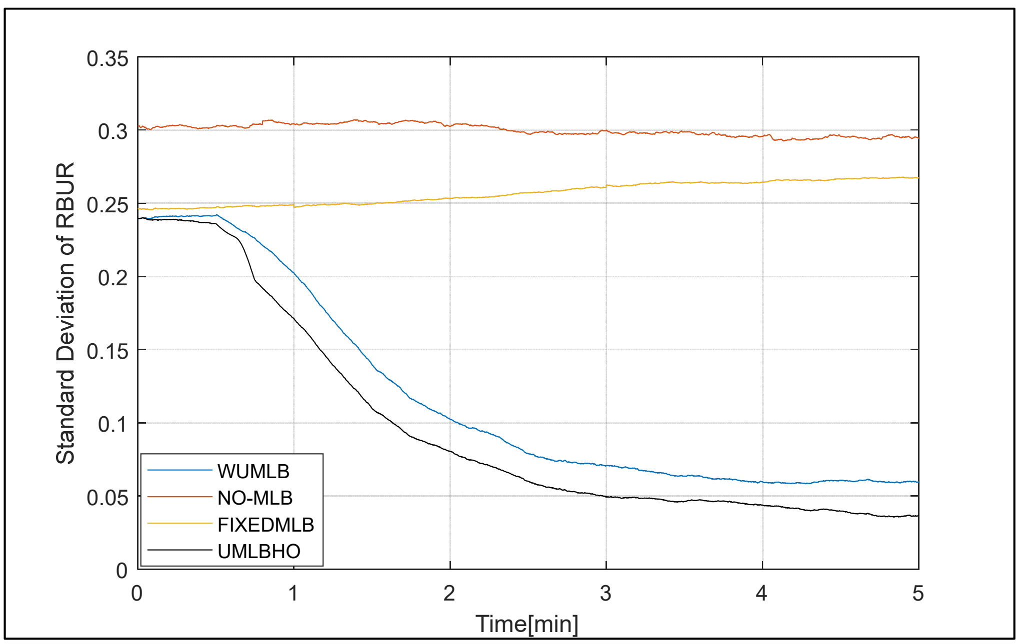



Figure 5.2 Standard deviation of RBUR among cells in the network

Moreover, we evaluated network performance in terms of the UE average data rate. Figure 5.3 shows the average UE rate for multiple MLB approaches. The proposed UMLB-HO algorithm provides 53% of UEs with an average data rate of more than 1Mbps. With the WUMLB algorithm, 37% of UEs with an average data rate of more than 1Mbps. On the other hand, approximately 3% of UEs can have an average data rate of 1Mbps when adopting No-MLB. While 10% of UEs can have an average data rate of 1Mbps when adopting Fixed-MLB algorithms.

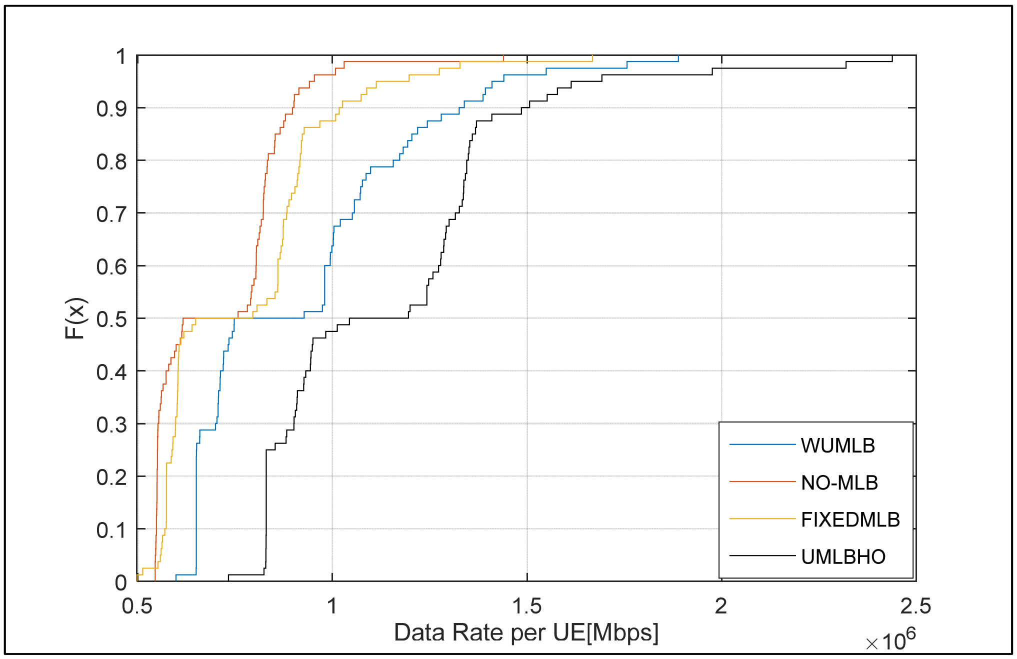



Figure 5.3 Average UE data rate [Mbps]

Finally, we evaluate the number of handovers that were performed by MLB algorithms. It is obvious that WUMLB does not categorize the UEs (Very slow, slow and fast) but for the sake of performance evaluation, we track the UEs throughout the simulation and detect its type (Very slow, slow and fast). Thus, we observe the handovers performed by each UE according to the WUMLB procedure. Consequently, we compare the UMLB-HO with the WUMLB and evaluate the performance. Figure 5.4 shows that the proposed UMLB-HO algorithm reduces the number of normal and forced handovers by 68.05%, 41.91% for Fixed-MLB, and WUMLB, respectively.

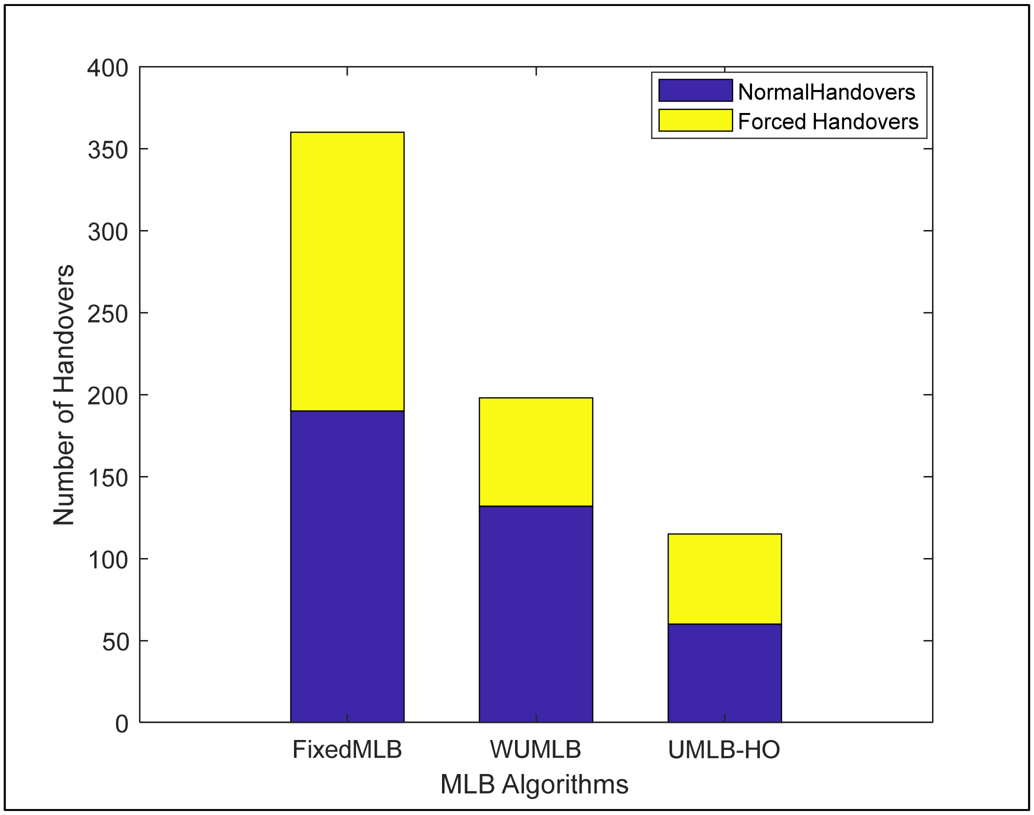



Figure 5.4 Total Number of Handovers

Figure 5.5 illustrates the normal and forced handovers classifications presented in Figure 5.4. We can see that the WUMLB algorithm causes 132 normal handovers which are equivalent to 60 A3-based handovers, 36 Frequent Fast Non-edge UEs' handovers (FF Non-Edge) and 30 Frequent Fast Edge UEs' handovers (FF Edge). While the WUMLB-HO algorithm causes only 60 normal handovers which are A3-based handovers, however; it avoids the frequent handovers (FF Non-Edge and FF Edge handovers).

Therefore, the UMLB-HO reduces the normal handover by 54.54% compared to the WUMLB. On the other hand, the WUMLB algorithm causes 66 forced handovers which are caused by 40 slow UEs' handovers, 6 Fast edge UEs' handovers and 20 very slow Edge UEs' handovers. While the WUMLB-HO algorithm causes only 55 forced which are equivalent to 40 slow UEs' handovers, 6 Fast edge UEs' handovers and 9 fast non-Edge UEs' handovers, however; it avoids the very- slow UEs' handovers. Therefore, the UMLB-HO reduces the forced handovers by 16.66% compared to the WUMLB.

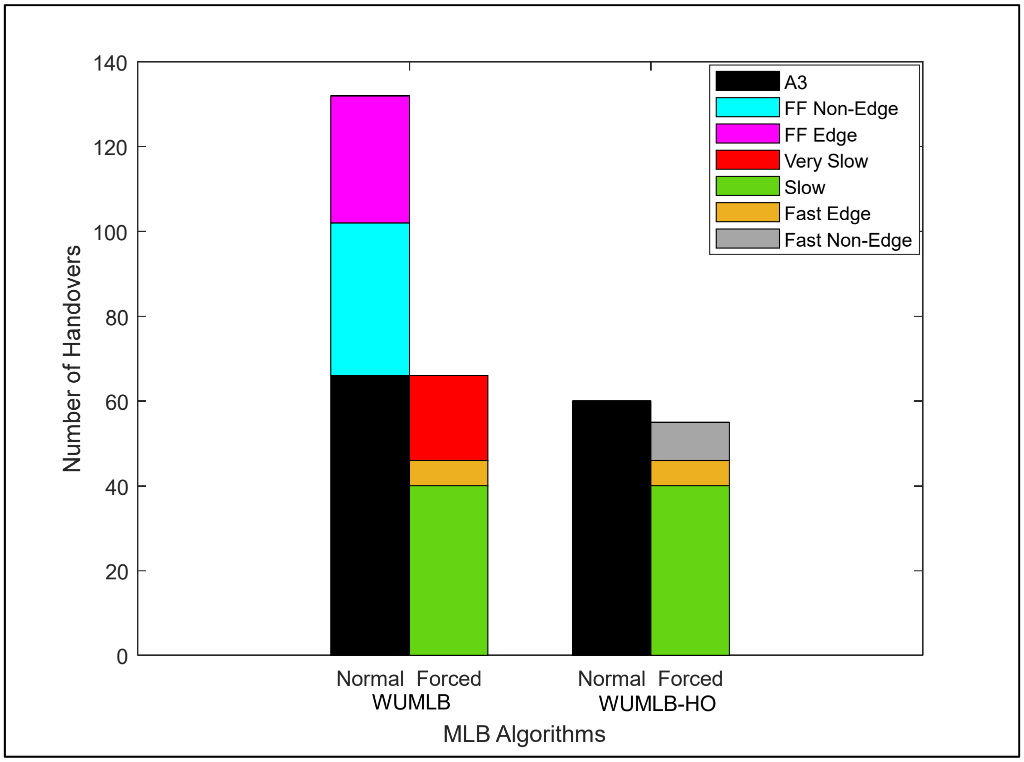



Figure 5.5 Number of Handover Types

## 5.6 Conclusion

In this chapter, we introduced a Utility-based Mobility Load Balancing algorithm with Handover minimization (UMLB-HO) by considering not only the edge UEs, but also the non-edge UEs for overloaded small cells during the load balancing process. The main objective for any overloaded small cell during the MLB is to determine the fast-moving non-edge UE and transfer them to the underloaded Macro cell. Moreover, the small cell will determine the fast-, slow or very slow-moving edge-UEs during the MLB and transfer them to either under loaded neighbour small cell or Macro cell. We defined four important terms to specify whether the UE is fast-, slow or very slow-moving in order to determine the best handover decisions for each UE. If the UE is a non-edge fast-moving, UE is going to be handed over to the Macro cell instead of small cells to avoid the unnecessary frequent handovers among the small cells. However, if the UE is an edge fast-moving, the UE will be handed over to the macro cell to avoid unnecessary frequent handovers. Moreover, if the UE is an edge slow-moving, he will be handed over to a neighbour small cell. While if the UE is a very slow-moving edge UE, he/she will not be handed over because his session will end before he leaves the serving cell, thus unnecessary handover is avoided. As a result, the load balancing is done with the minimum number of handovers. Simulation results show that the proposed UMLB-HO algorithm has the minimum number of handovers for a minimum standard deviation with an enhanced level of throughput.

## CONCLUSIONS AND FUTURE WORKS

### Conclusions

Enhanced throughput and guaranteed coverage extension can be provided by using mobile relaying techniques in LTE-A system Networks. 3GPP started supporting mobile relay in LTE-Networks (Rel.12). In this thesis, we present the carrier aggregation-supported mobile relay for railway LTE-A networks. The performance of the novel scheme is evaluated in terms of average user throughput, packet loss and outage probability. Results show that better performance can be accomplished by applying dual-backhaul links with carrier aggregation scheme.

This thesis also investigates the user association problem in the downlink LTA-A network with mobile relay nodes. However, this is a complicated process for large-scale heterogeneous networks. Thus, a matching game-based user association scheme was introduced for solving the assignment problem to increase the number of admitted users in the system under resource constraints and QoS requirements. The user association algorithm is called the Chance-Based Deferred Acceptance Matching (CBDAM) Algorithm. Particularly, the new proposed approach maximizes the number of users admitted in the system and the total system throughput by assigning each user (macro or vehicular) to the stations based on negotiations between eNBs and users. With mobility existence, the handover failure rate is significantly decreased for the vehicular users.

After that, we introduced the UMLB algorithm that balances the load across a small-cell network by considering the operator utility and the user utility for the handover process. The operator utility is calculated for each potential handover based on the load of the neighbouring small cells. Whereas, the user utility calculation is based on the sigmoid function by considering different criteria. Also, we presented a new term named load balancing efficiency factor (LBEF). The LBEF considers a load of neighbouring cells and the edge-UEs for each overloaded cell. This factor specifies the sequence of overloaded cells for the UMLB algorithm operation. The simulation results show that the UMLB minimizes standard deviation with a higher average-UE data rate when compared to existing load balancing algorithms. Therefore, a well-balanced network is achieved.

Finally, we enhanced the UMLB by minimizing the number of handovers performed in the small cell networks, which is the UMLB-HO algorithm. The transfer decision is based on the classification of the candidate UEs within an overloaded cell. Some hangovers are completely avoided, others, are made to the macro cell to minimize the frequent handovers. The simulation results show that the UMLB-HO minimizes standard deviation with a minimum number of handovers compared to UMLB algorithm.

**Future Work**

We believe that there will always be a room for improvement and things to be done in the future, some of the potential future research directions are:

- A comprehensive MLB algorithm that considers all types of readio access networks is a massive challenge to address due to the different parameters and constraints pertaint to each technology.
- Underutilized cells may be turned off to save power and thus the environment, however; switching On/Off cells needs the cooperation from all nodes in the network to guarantee UEs the required QoS and connectivity.
- As this thesis concentrates on mobility load balancing in heterogeneous networks, transferring load from macro to small cells and adapting overload threshold to macrocells should be considered. Originally, small cells were found to increase the capacity or expand the coverage of Macro cells.
- Estimate the classification accuracy of UE movements and introduce appropriate modifications to enhance the performance of the proposed algorithm.

## APPENDIX

## LIST OF PUBLICATIONS

**Journals:**

**Accepted and Published**

K. M. Addali, S. Y. Bani Melhem, Y. Khamayseh, Z. Zhang and M. Kadoch, "Dynamic Mobility Load Balancing for 5G Small-Cell Networks Based on Utility Functions," in *IEEE Access*, vol. 7, pp. 126998-127011, 2019.

**Conferences:**

**Accepted and Published**

K. M. Addali, A. BenMimoune, F. A. Khasawneh, A. M. Saied and M. Kadoch, "Dual-Backhaul Links in LTE-A Mobile Relay System for High-Speed Railways," *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, Vienna, 2016, pp. 98-102.

K. M. Addali and M. Kadoch, "Matching Game-Based User Association Approach for LTE-A Mobile Relay Systems," *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, Quebec City, QC, 2018, pp. 1-6.

K. Addali and M. Kadoch, "Enhanced Mobility Load Balancing Algorithm for 5G Small Cell Networks," *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada, 2019, pp. 1-5.

**Accepted**

K. Addali, Zhengwei Chang, Jizhao Lu, Rongke Liu, and M. Kadoch, " Mobility Load Balancing with Handover Minimization for 5G Small Cell Networks" *International conference on Wireless Communications & Mobile Computing (IWCMC 2020)*, Limassol, Cyprus.

## LIST OF BIBLIOGRAPHICAL REFERENCES

3GPP. (2012a). Evolved Universal Terrestrial Radio Access (E-UTRA); FDD Repeater Radio Transmission and Reception, document TS 36.106, .

3GPP. (2012b). TR 36.836 Evolved Universal Terrestrial Radio Access (EUTRA); Study on mobile relay.

3GPP. (2014). Universal Mobile Telecommunications System (UMTS); LTE; Telecommunication Management; Self-Organizing Networks (SON); Concepts and requirements, TS 32.500, .

3GPP. (2014.). Universal Mobile Telecommunications System (UMTS); LTE; Telecommunication Management; Self-Organizing Networks (SON); Concepts and requirements, .

3GPP. (2016). Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP), document TS 36.423, .

3GPP. (2019). Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S2 Application Protocol (S1AP), document TS 36.413, .

Access, E. U. T. R. (2009). Physical Layer Procedures, 3GPP Technical Specification TS 36.213: Rev.

Addali, K. M., BenMimoune, A., Khasawneh, F. A., Saied, A. M., & Kadoch, M. (2016). Dual-Backhaul Links in LTE-A Mobile Relay System for High-Speed Railways. Dans *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)* (pp. 98-102). IEEE.

Albasheir, S. K. (2016). *Enhanced resource utilization in the evolved packet core network for 4G/LTE mobile broadband services* (École de technologie supérieure).

Alexandris, K., Nikaein, N., Knopp, R., & Bonnet, C. (2016). Analyzing x2 handover in lte/lte-a. Dans *2016 14th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt)* (pp. 1-7). IEEE.

Alsharoa, A., Ghazzai, H., Yaacoub, E., & Alouini, M.-S. (2014). Energy-efficient two-hop LTE resource allocation in high speed trains with moving relays. Dans *2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)* (pp. 528-533). IEEE.

Andersen, J. B., Rappaport, T. S., & Yoshida, S. (1995). Propagation measurements and models for wireless communications channels. *IEEE Communications Magazine, 33*(1), 42-49.

Atat, R., Yaacoub, E., Alouini, M.-S., & Abu-Dayya, A. (2012). Heterogeneous LTE/802.11 a mobile relays for data rate enhancement and energy-efficiency in high speed trains. Dans *2012 IEEE Globecom Workshops* (pp. 421-425). IEEE.

Bian, Y. Q., & Rao, D. (2014). Small cells big opportunities. *Global Business Consulting. Huawei Technologies Co., Ltd.*

Bing, H., He, C., & Jiang, L. (2003). Performance analysis of vertical handover in a UMTS-WLAN integrated network. Dans *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003.* (Vol. 1, pp. 187-191). IEEE.

Cho, H., Park, J., Ko, W., Lim, K., & Kim, W. (2005). A study on the MCHO method in Hard handover and Soft handover between WLAN and CDMA. Dans *2005 Digest of Technical Papers. International Conference on Consumer Electronics, 2005. ICCE.* (pp. 391-392). IEEE.

Chou, C. M., & Huang, C. (2006). Dynamic vertical handover control algorithm for WLAN and UMTS. Dans *IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006.* (Vol. 1, pp. 606-610). IEEE.

Cisco. (2019). Cisco visual networking index: Global mobile data trafc forecast update 2017-2022. doi: 1486680503328360,

Dahlman, E., Parkvall, S., & Skold, J. (2013). *4G: LTE/LTE-advanced for mobile broadband*. Academic press.

Damnjanovic, A., Montojo, J., Wei, Y., Ji, T., Luo, T., Vajapeyam, M., . . . Malladi, D. (2011). A survey on 3gpp heterogeneous networks. Wireless Comm: IEEE.

Diederich, J., & Zitterbart, M. (2005). Handoff prioritization schemes using early blocking. *IEEE Communications surveys & tutorials, 7*(2), 26-45.

*Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN),* . Repéré à http://www.3gpp.org/

Fedrizzi, R., Goratti, L., Rasheed, T., & Kandeepan, S. (2016). A heuristic approach to mobility robustness in 4G LTE public safety networks. Dans *2016 IEEE Wireless Communications and Networking Conference* (pp. 1-6). IEEE.

Feng, S., & Seidel, E. (2008). Self-organizing networks (SON) in 3GPP long term evolution. *Nomor Research GmbH, Munich, Germany, 20*.

Forum, S. C. (2014). Small cells, what's the big idea? *, Tech. Rep. SCF030*

Hasan, M. M., Kwon, S., & Na, J.-H. (2018). Adaptive mobility load balancing algorithm for LTE small-cell networks. *IEEE Transactions on wireless Communications, 17*(4), 2205-2217.

Hoadley, J., & Maveddat, P. (2012). Enabling small cell deployment with HetNet. *IEEE Wireless Communications, 19*(2), 4-5.

Huang, Z., Liu, J., Shen, Q., Wu, J., & Gan, X. (2015). A threshold-based multi-traffic load balance mechanism in LTE-A networks. Dans *2015 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1273-1278). IEEE.

Kwan, R., Arnott, R., Paterson, R., Trivisonno, R., & Kubota, M. (2010). On mobility load balancing for LTE systems. Dans *2010 IEEE 72nd Vehicular Technology Conference-Fall* (pp. 1-5). IEEE.

Li, W., Zhang, C., Duan, X., Jia, S., Liu, Y., & Zhang, L. (2012). Performance evaluation and analysis on group mobility of mobile relay for LTE advanced system. Dans *2012 IEEE Vehicular Technology Conference (VTC Fall)* (pp. 1-5). IEEE.

Lobinger, A., Stefanski, S., Jansen, T., & Balan, I. (2010). Load balancing in downlink LTE self-optimizing networks. Dans *2010 IEEE 71st Vehicular Technology Conference* (pp. 1-5). IEEE.

Lobinger, A., Stefanski, S., Jansen, T., & Balan, I. (2011). Coordinating handover parameter optimization and load balancing in LTE self-optimizing networks. Dans *2011 IEEE 73rd vehicular technology conference (VTC Spring)* (pp. 1-5). IEEE.

Luo, W., Zhang, R., & Fang, X. (2012). A CoMP soft handover scheme for LTE systems in high speed railway. *EURASIP Journal on Wireless Communications and Networking, 2012*(1), 1-9.

Lv, J., Ma, Y., & Yoshizawa, S. (2008). Intelligent Seamless Vertical Handoff Algorithm for the next generation wireless networks. Dans *Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications* (pp. 23). ICST (Institute for Computer Sciences, Social-Informatics and ….

Moghaddam, M., & Nof, S. Y. (2017). *Best matching theory & applications*. Springer.

Nasser, N., Hasswa, A., & Hassanein, H. (2006). Handoffs in fourth generation heterogeneous networks. *IEEE Communications Magazine, 44*(10), 96-103.

Nelakuditi, S., Harinath, R. R., Rayadurgam, S., & Zhang, Z.-L. (1999). Revenue-based call admission control for wireless cellular networks. Dans *1999 IEEE International Conference on Personal Wireless Communications (Cat. No. 99TH8366)* (pp. 486-490). IEEE.

Nguyen-Vuong, Q.-T., Ghamri-Doudane, Y., & Agoulmine, N. (2008). On utility models for access network selection in wireless heterogeneous networks. Dans *NOMS 2008-2008 IEEE Network Operations and Management Symposium* (pp. 144-151). IEEE.

Nicopolitidis, P., Obaidat, M. S., Papadimitriou, G. I., & Pomportsis, A. S. (2003). *Wireless networks*. Wiley Online Library.

NTT, D. (2010). R1-103264: Performance of eICIC with control channel coverage limitation. Dans *3GPP TSG RAN WG1, Meeting# 61, Montreal, Canada, May 2010*.

Oh, S., Kim, H., & Kim, Y. (2018). User mobility impacts to mobility load balancing for self-organizing network over LTE system. Dans *2018 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET)* (pp. 1082-1086). IEEE.

Oh, S., Kim, H., Na, J., Kim, Y., & Kwon, S. (2016). Mobility load balancing enhancement for self-organizing network over lte system. Dans *Internet of Things, Smart Spaces, and Next Generation Networks and Systems* (pp. 205-216). Springer.

Pan, M.-S., Lin, T.-M., & Chen, W.-T. (2014). An enhanced handover scheme for mobile relays in LTE-A high-speed rail networks. *IEEE Transactions on Vehicular Technology, 64*(2), 743-756.

Parkvall, S., Furuskar, A., & Dahlman, E. (2011). Evolution of LTE toward IMT-advanced. *IEEE Communications Magazine, 49*(2), 84-91.

Qualcomm. (2010). LTE Advanced: Heterogeneous Networks, .

Qualcomm. (2019). Small Cells With UltraSON. Repéré à https://www.qualcomm.com/media/documents/files/small-cells-and-ultrason-presentation.pdf

Rajpoot, P., & Dwivedi, P. (2018). Optimized and load balanced clustering for wireless sensor networks to increase the lifetime of WSN using MADM approaches. *Wireless Networks*, 1-37.

Saad, W., Han, Z., Zheng, R., Debbah, M., & Poor, H. V. (2014). A college admissions game for uplink user association in wireless small cell networks. Dans *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (pp. 1096-1104). IEEE.

Schmelz, L. C., Amirijoo, M., Eisenblaetter, A., Litjens, R., Neuland, M., & Turk, J. (2011). A coordination framework for self-organisation in LTE networks. Dans *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops* (pp. 193-200). IEEE.

Tanghe, E., Joseph, W., Verloock, L., & Martens, L. (2008). Evaluation of vehicle penetration loss at wireless communication frequencies. *IEEE Transactions on Vehicular Technology, 57*(4), 2036-2041.

Teyeb, O., Van Phan, V., Raaf, B., & Redana, S. (2009). Dynamic relaying in 3GPP LTE-Advanced networks. *EURASIP Journal on Wireless Communications and Networking, 2009*, 1-11.

Trabelsi, N., Chen, C. S., El Azouzi, R., Roullet, L., & Altman, E. (2017). User association and resource allocation optimization in LTE cellular networks. *IEEE Transactions on Network and Service Management, 14*(2), 429-440.

Van Quang, B., Prasad, R. V., & Niemegeers, I. (2010). A survey on handoffs—Lessons for 60 GHz based wireless systems. *IEEE Communications surveys & tutorials, 14*(1), 64-86.

Vesterstrom, J., & Thomsen, R. (2004). A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. Dans *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)* (Vol. 2, pp. 1980-1987). IEEE.

Vu, T.-T., Decreusefond, L., & Martins, P. (2014). An analytical model for evaluating outage and handover probability of cellular wireless networks. *Wireless Personal Communications, 74*(4), 1117-1127.

Yamamoto, T., Komine, T., & Konishi, S. (2012). Mobility load balancing scheme based on cell reselection. *Cell, 7*, 7UEs.

Yang, K., Gondal, I., Qiu, B., & Dooley, L. S. (2007). Combined SINR based vertical handoff algorithm for next generation heterogeneous wireless networks. Dans *IEEE GLOBECOM 2007-IEEE Global Telecommunications Conference* (pp. 4483-4487). IEEE.

Yu, F., & Leung, V. (2002). Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks. *Computer Networks, 38*(5), 577-589.

Zhang, X., & Wang, P. (2014). Group-based collaborative handover in LTE macro-femtocell network.

Zhao, H., Huang, R., Zhang, J., & Fang, Y. (2011). Handoff for wireless networks with mobile relay stations. Dans *2011 IEEE Wireless Communications and Networking Conference* (pp. 826-831). IEEE.

Zhu, F., & McNair, J. (2006). Multiservice vertical handoff decision algorithms. *EURASIP Journal on Wireless Communications and Networking, 2006*(1), 025861.

Zia, N., & Mitschele-Thiel, A. (2013). Self-organized neighborhood mobility load balancing for LTE networks. Dans *2013 IFIP Wireless Days (WD)* (pp. 1-6). IEEE.