12-20-2017 11:00 AM

# Intelligent Advancements in Location Management and C-RAN Power-Aware Resource Allocation

Emad Ali Aqeeli
*The University of Western Ontario*

Supervisor
Abdallah Shami
*The University of Western Ontario*

Graduate Program in Electrical and Computer Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

# Abstract

The evolving of cellular networks within the last decade continues to focus on delivering a robust and reliable means to cope with the increasing number of users and demanded capacity. Recent advancements of cellular networks such as Long-Term Evolution (LTE) and LTE-advanced offer a remarkable high bandwidth connectivity delivered to the users. Signalling overhead is one of the vital issues that impact the cellular behavior. Causing a significant load in the core network hence effecting the cellular network reliability. Moreover, the signaling overhead decreases the Quality of Experience (QoE) of users.

The first topic of the thesis attempts to reduce the signaling overhead by developing intelligent location management techniques that minimize paging and Tracking Area Update (TAU) signals. Consequently, the corresponding optimization problems are formulated. Furthermore, several techniques and heuristic algorithms are implemented to solve the formulated problems.

Additionally, network scalability has become a challenging aspect that has been hindered by the current network architecture. As a result, Cloud Radio Access Networks (C-RANs) have been introduced as a new trend in wireless technologies to address this challenge. C-RAN architecture consists of: Remote Radio Head (RRH), Baseband Unit (BBU), and the optical network connecting them. However, RRH-to-BBU resource allocation can cause a significant downgrade in efficiency, particularly the allocation of the computational resources in the BBU pool to densely deployed small cells. This causes a vast increase in the power consumption and wasteful resources.

Therefore, the second topic of the thesis discusses C-RAN infrastructure, particularly where a pool of BBUs are gathered to process the computational resources. We argue that there is a need of optimizing the processing capacity in order to minimize the power consumption and increase the overall system efficiency. Consequently, the optimal allocation of computational resources between the RRHs and BBUs is modeled. Furthermore, in order to get an optimal RRH-to-BBU allocation, it is essential to have an optimal physical resource allocation for users to determine the required computational resources. For this purpose, an optimization problem that models the assignment of resources at these two levels (from physical resources to users and from RRHs to BBUs) is formulated.

**Keywords:** LTE, Location Management, Tracking Area (TA), Tracking Area List (TAL), MME pooling, SON, C-Ran, LTE, RRH, BBU, Resource allocation, Power consumption, Computer resource allocation.

# Acknowledgements

I first thanks “Allah” who flourishes my life with countless gifts and enlightens my soul with wisdom and patience. Gifts that are nothing a like, the wonderful souls that exist in my life such as my family especially my Parents “Ali and Saliha” who devoted their time and youth to help me become what I am now. I will always be grateful and no words or deeds are enough to repay them.

Special thanks to my kind wife “Miad” who stood and supported me in every step, her patience, prayers and the unforgettable smile that helped to ease this journey. Another thanks to my beloved son “Wisam”, my soul’s light, to his laughter, love, and all the great achievements that he managed to accomplish. In addition, a special thanks to the coming soul.

My deepest gratitude and respect goes to my brothers and sisters:“Mohamed, Aziza, Abdulrahman, Najat and Essam” who supported me in uncountable ways and prayed for me to become a successful person like them.

A special appreciation and respect to my supervisor Professor Abdallah Shami, who helped me through my degree by his continuous support, guidance and encouragement to always aim to the best.

A special thanks to all my colleges and friends who also become as a family that always were there for support: Abdallah Moubayed, Abdulfattah Noorwali, Abdulbaset, Anas Saci, Ammer, Hassan & Manar Hawilo, Ibrahim. Khaleel Sunba, Karim Hammad, Khaled Al Hazmi, Khalid Abo zain, Mohamed Kalil, Mohamed Abu Sharkh, Mohamed Noor, Mohamed Hussein, Monagi Alkinani, and all my other friends.

At the end, I send my thanks to the Royal Commission for Yanbu for giving me the great opportunity to complete my studies and a special thanks to Dr.Abdulkarim Al-Alwani who helped me with his advise and wisdom. Another thanks goes to the Saudi Cultural Bureau in Canada, to each and everyone of them.

***Emad Aqeeli***

*To the memory of my grandmother "Haifa Aqeeli"*

# Table of Contents

# List of Tables

# List of Figures

**Section** **Page**

# List of Abbreviations

| | |
|---|---|
| 1G | First Generation |
| 2G | Second Generation |
| ABC | Artificial Bee Colony |
| BBU | Baseband Unit |
| BIP | Binary Integer Programming |
| CAPEX | Capital Expenditures |
| C-RAN | Cloud Radio Access Network |
| CCI | Co-Channel Interference |
| DL | Downlink |
| EPC | Evolved Packet Core |
| FFD | First Fit Decreasing |
| GSA | Gravitational Search Algorithm |
| GPRS | General Packet Radio Service |
| HSS | Home Subscriber Server |
| IP | Internet Protocol |
| LA | Location Area |
| LP | Linear Problem |
| LTE | Long-Term Evolution |
| MINLP | Mixed-Integer Nonlinear Program |
| MME | Mobility Management Entity |
| MPOTS | Million Operations Per Time Slot |
| OPEX | Operating Expenditures |
| PCRF | Policy and Charging Rules Function |

| | |
|---|---|
| PCEF | Policy and Charging Enforcement Function |
| PDN | Packet Data NetworK |
| P-GW | Packet Data Network Gateway |
| PHY | Physical Layer |
| PSO | Particle Swarm Optimization |
| QoE | Quality of Experience |
| QoS | Quality of Service |
| RA | Routing Area |
| RB | Resource Block |
| RRH | Remote Radio Head |
| RNC | Radio Network Controller |
| RSD | Relative Standard Deviation |
| S-GW | Serving Gateway |
| SNR | Signal to Noise Ratio |
| SON | Self-Organizing Network |
| TA | Tracking Area |
| TAL | Tracking Area List |
| TAU | Tracking Area Updates |
| UE | USER EQUIPMENT |
| UP | Uplink |
| VBS | Virtual Base Stations |

# List of Symbols

| | |
|---|---|
| $a$ | agent acceleration |
| $B$ | RB bandwidth (typically 180 KHz) |
| $B_j$ | $\begin{cases} 1, & \text{if BBU } j \text{ is used,} \\ 0, & \text{otherwise} \end{cases}$ |
| $C_\rho$ | Paging cost of particular user equipment |
| $C_{T\rho}(i)$ | Total paging cost in cell $i$ |
| $Cu\rho(i)$ | Total paging and TAU overhead in cell $i$ |
| $C_{Tu}(i)$ | Total signaling cost of TAU in cell $i$ |
| $C_u$ | TAU cost of UE moving from one list to another |
| $\mathcal{F}$ | Gravitational force |
| $G$ | Gravitational constant |
| $G^{rb}_{(i,k)}$ | Channel condition between RRH $i$ and user $k$ on RB $rb$ |
| $HX^{lst}(i)$ | Inter-list handover rate of users in cell $i$ |
| $L$ | Total number of lists |
| $L^j$ | Available computing resources in each BBU $j$ |
| $L^j_i$ | Total computing resources given by BBU $j$ to cell $i$ |
| $L^{tot}_J$ | Total computing resources used by BBU $j$ |
| $lst$ | Individual list where $lst$ is a subset of $L$ |
| $H_{ij}$ | The probability that a user moves from cell $i$ to cell $j$ |
| $\mathcal{J}$ | Total cost function |
| $J_1$ | Cost function 1 to be minimized |
| $J_2$ | Cost function 2 to be minimized |
| $\|K\|$ | Total number of users |
| $\kappa$ | Maximum number of TAs that are assigned to list $lst$ |
| $\|M\|$ | Total number of BBUs |
| $\mathcal{M}$ | Mass of agent |
| $\|N\|$ | Total number of RRHs/Cells |
| $\mathcal{N}$ | Population size |

| | |
|---|---|
| $N_0$ | Noise |
| $O_k^i$ | $\begin{cases} 1, & \text{if user } k \text{ is served by cell } i, \\ 0, & \text{otherwise} \end{cases}$ |
| $O_{lst}^M$ | $\begin{cases} 1, & \text{if list } lst \text{ belongs to MME } M, \\ 0, & \text{otherwise} \end{cases}$ |
| $P$ | Probability of food source |
| $P_{pool}$ | Power consumption of the BBU pool |
| $P_i$ | Power consumption of RRH/Cell $i$ |
| $P_j^{tot}$ | Total power consumption of an individual BBU $j$ |
| $P^{TOT}$ | Total power consumption |
| $P_{st}^{pool}$ | Static power consumption of an individual pool $s$ |
| $P_{st}^j$ | Static power of each BBU obtained by well-known procedures |
| $\mathcal{P}$ | Number of parameters to be optimized |
| $\rho$ | Paging arrival rate |
| $\mathcal{R}$ | Euclidean distance |
| $\|RB\|$ | Total number of RBs |
| $r_{k,th}$ | Minimum data rate requirement of user $k$ |
| $\sigma_i^{lst}$ | Percentage use of each list $lst$ in cell $i$ |
| $UE_i$ | Total number of UEs served by cell $i$ |
| $v$ | velocity of particle or agent |
| $\omega$ | Cost of MME relocation during handover |
| $X_{ij}^{lst}$ | $\begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ belong to } lst, \\ 0, & \text{otherwise} \end{cases}$ |
| $x$ | position of particle, food source or agent |
| $x_k^{rb}$ | $\begin{cases} 1, & \text{if user } k \text{ is allocated RB } rb, \\ 0, & \text{otherwise} \end{cases}$ |
| $Y_i^{TA}$ | $\begin{cases} 1, & \text{if node } i \text{ belongs to } TA, \\ 0, & \text{otherwise} \end{cases}$ |

# Chapter 1
# Introduction

## 1.1 Historical Overview of Cellular Networks

Since early eighties, mobile network has gained fair attention from both industry and academia. It all started when the First Generation (1G) of the analog cellular networks has been introduced as the Advanced Mobile Phone System (AMPS). After ten years, the Second Generation (2G) was launched to support mobile phones and limited data connection in the (2.5G) extension. With the support of General Packet Radio Service (GPRS), the (2.5G) networks use the circuit switching for the voice and the packet switching for the data transmission. The evolution continued with introducing the third generation that enables faster speeds with better efficiency and quality of service. The fourth generation then has introduced two standards called WiMAX and Long Term Evolution (LTE) enhances the capability of packet switching to provide the users an astonishing performance. LTE has gained more attention than WiMAX, as it supports higher speeds, better performance, and scalable bandwidth.

## 1.2 Motivation

Wireless technologies continue to be challenged with the momentous traffic that affects bandwidth. Moreover, the continuous proliferation of hand-held devices and their applications generate a surge of signaling traffic with a greater percentage than the user data traffic. The signaling traffic initiated each time there is a transmission or reception of packet streams between the UEs and mobile network despite the actual size of data traffic. Nokia Siemens Networks has predicted the increase of signaling in the coming years will be up to 50% faster than the growth of data traffic. In addition, LTE experiences more signaling overhead than other 3G technologies due to its flat IP architecture that do not contain a medium entity such as Radio Network Controller (RNC) between the base station and the core network [6]. For instance the average signaling overhead is 42% more than HSPA per subscriber.

Paging and Tracking Area Updates (TAU) have the most significant signaling impact on the Evolved Packet Core (EPC), specifically on the Mobility Management Entity (MME). Paging and TAU are essential functions for UE location management. They are used to track the user's location and provide constant updates to the EPC. Paging and TAU are defined as follows:

1. Paging: messages sent by the MME to locate a particular UE in a Tracking Area (TA).

2. Tracking area update: messages sent by UEs to the MME when they move from one tracking area to another.

To identify the UE location, the LTE core network pages the latest tracking area that the UE was registered to. The paging signal is received by all of the cells that reside in the same tracking area. Additionally, the UE will update the core network by sending a tracking area update (TAU) signal once it moves from one tracking area to another.

Tracking Area (TA) is the area in which the MME can locate a specific user within a defined set of cells. This technique is used in LTE and was originally inherited from previous 2G and 3G technologies. However, TA has a number of limitations that have led to the introduction of the new concept of Tracking Area List (TAL), whereby several TAs are grouped into a single TAL. TAL has the same functionality as TA but with the added flexibility of a set of TAs within the TAL. Thus, TAL can alleviate the signaling load due to triggering TAU each time a UE moves from one cell to another.

Several studies have investigated cell-to-TA/TAL assignment problem as they used various techniques to optimize the system by minimizing the signaling overhead. At present, none have investigated TA/TAL-to-MME assignment, which has the same importance as investigating cell-to-TA/TAL assignment. TA/TAL-to-MME allocation is vital for minimizing the signaling overhead gained from the TAU-to-core network, especially in the MME. The MME receives a large amount of signaling traffic resulting from several UEs, each making different requests. As a result, it is more valuable to consider a mechanism to intelligently relate and distribute the signaling load to the MME during cell-to-TA/TAL construction. LTE has introduced the concept of MME pooling, in which the number of clustered MMEs can perform the mobility management function as one MME. Consequently, this decreases the amount of signaling incurred during various procedures such as MME relocation. Ultimately, determining the type of MME pooling versus the TA/TAL allocation will give more accurate control and allow optimization of the signaling overhead resulting from paging and TAU.

In Release 8, 3GPP introduced the concept of a Self-Organizing Network (SON) that provides a methodology for planning, managing, and optimizing mobile networks in order improve performance efficiency and system reliability. SON has been widely accepted in industry and academia [7, 8]. 3GPP has also released different use cases for LTE, offering self-optimizing and self-healing paradigms. In this context, adaptive TA list management can be used as an SON use case as in [9], which can further reduce the signaling load. Therefore, the second part of the thesis introduces different techniques that relate the TA list to the behavior of the mobile network would further optimize the signaling overhead. Thus, cell-to-TAL assignment can be engineered dynamically while the UE is in continuous movement. The system keeps analyzing the mobility pattern and continuously updates the TA assigned to the list. Thus, the frequency of TAU will be reduced significantly. In this part, the mobility pattern is obtained using a fluid flow model to estimate the handover correlation between cells.

Due to the complexity of the original model, the third part of the thesis uses the evolutionary artificial techniques in optimizing the location management techniques when applied in a large-scale network. Three algorithms are implemented namely; Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), and Gravitational Search Algorithm (GSA). To the best of our knowledge, the proposed study is the first to include these algorithms.

In recent years, PSO, ABC and GSA have been considered to be reliable and promising evolutionary algorithms for locating the global optima for a variety of optimization problems in different contexts with rapid rate of convergence as seen in different applications [2, 4, 10, 11]. The three selected algorithms belong to different heuristic search families such as PSO relies on swarming and nature behavior of birds [1], ABC stands on colony structure and the main role of three main groups within the colony [3] and GSA is based on gravity laws and mass interactions between particles [5]. The obtained simulation results illustrate the effectiveness of the proposed approach in finding minimum total signaling overhead resulted in paging and tracking area update. Moreover, the applied optimization techniques offer an efficient solution for large scaled problems that need fast convergence. Lastly, The results show that the power consumption of the user equipment hand-held devices is minimized as well.

In the last part of the thesis Cloud Radio Access Network (C-RAN) is considered for optimizing the data traffic. C-RAN has been introduced as a new paradigm that has succeeded in bringing forth a new era to the world of wireless communications. C-RAN was first introduced by China Mobile, one of the key wireless operators in China. The advantages of cloud computing have influenced researchers in both academia and industry to integrate the cloud paradigm into different applications. Featuring a simple yet clever solution that splits up conventional base stations into two independent entities, namely a Remote Radio Head (RRH) and a Baseband Unit (BBU), connected through an optical high-speed transport network. Cloud computing enables real-time centralized processing through a virtualized BBU pool. In addition, the cloud architecture provides several advantages for C-RAN in numerous aspects, such as reducing the total CAPEX and OPEX as well as providing flexibility by distributing the capacity of the system.

The separation of the computational resources from the RRH has resulted in a significant reduction of power and has increased spectral efficiency. Moreover, the RRHs are deployed as small cells that can be densely distributed in a way that causes minimum interference. Thus, the distance between the User Equipment (UE) and RRHs is minimized, allowing for a more stabilized throughput gain. In addition, the implementation of BBU pools help to alleviate the energy cost of transmission and reception between the RRH and the BBU.

However, the efficiency of C-RAN is heavily dependent on the processing resources available in the BBU pool. In other words, there is a tight correlation between the efficiency of the mapping of computing resources from BBU to RRH and the overall performance of the system. The dense deployment of small cell imposes the necessity of distributing and allocating the resources for RRHs in the BBU pool intelligently. In 2006, the consumed electricity of data centers was estimated as 1.5% of the total generated electricity which is similar to the annual consumption of around 5.8 million households [12]. Consequently, various studies have been conducted to shade the light on the importance of alleviating the power consumption of cloud data centers. In [13], authors presented various methodologies for resource management in cloud data centers in order to increase their energy efficiency. A workload predictions through machine learning and stochastic theory [14] is an example of different study proposed an energy-aware framework for cloud data centers. The study has concluded that the utilization of Virtual Machines (VMs) is a key aspect in the power consumption of cloud data centers where a less utilized VMs leads to significant increase of VMs' requests. As a result, more physical resources will be accommodated. Therefore, resource scheduling between RRH and BBU is essential for efficient and reliable C-RAN implementation. In order to quantify these resources, the system calculates the computing requirements of each RRH and accordingly distributes the resources available within the BBU pools.

On the other hand, the computing requirements of each RRH are related to the scheduling of the physical resources to the users such that the Quality of Service (QoS) is satisfied. Because of this, the system has two levels of scheduling, each of which carries the same importance. The system first ensures user satisfaction by suitably distributing the physical resources and then schedules the baseband processing requirements simultaneously to all the BBU pools. Therefore, it is vital to optimize the scheduling process at each level. Moreover, optimizing computing resources in cloud systems is important because of the increased demand caused by intensive applications. In addition, network operators seek to minimize the cost of expanding resources while also having adequate computing resources available for instant processing.

## 1.3 Evolved Packet Core (EPC) Architecture

EPC offers the core functionality for LTE technology as an end to end IP based core network. EPC was introduced in 3GPP Release 8; it has the ability of transferring packets from the user equipment to the other end such as internet, VoIP, and etc., in IP based architecture. It also supports other technologies like 2G and 3G.

### 1.3.1 EPC Key Roles

- Session Management. The main function of EPC is that it handles the establishment of UE session by assigning bearers with specific quality of service requirements.

- Security and Privacy. Considered to be one of the essential functions of EPC. EPC is responsible for authenticating and providing encryption to UE for privacy.

- Mobility Management. EPC has the ability of tracking users among their movements between different cells as well as managing handovers and location registration.

- Policy and Charging Control. It has the role of categorizing users' policies and enforcing the quality of service accordingly.

### 1.3.2 EPC Essential Components

The essential elements of EPC are depicted in Figure 1.1, and explained as follows:

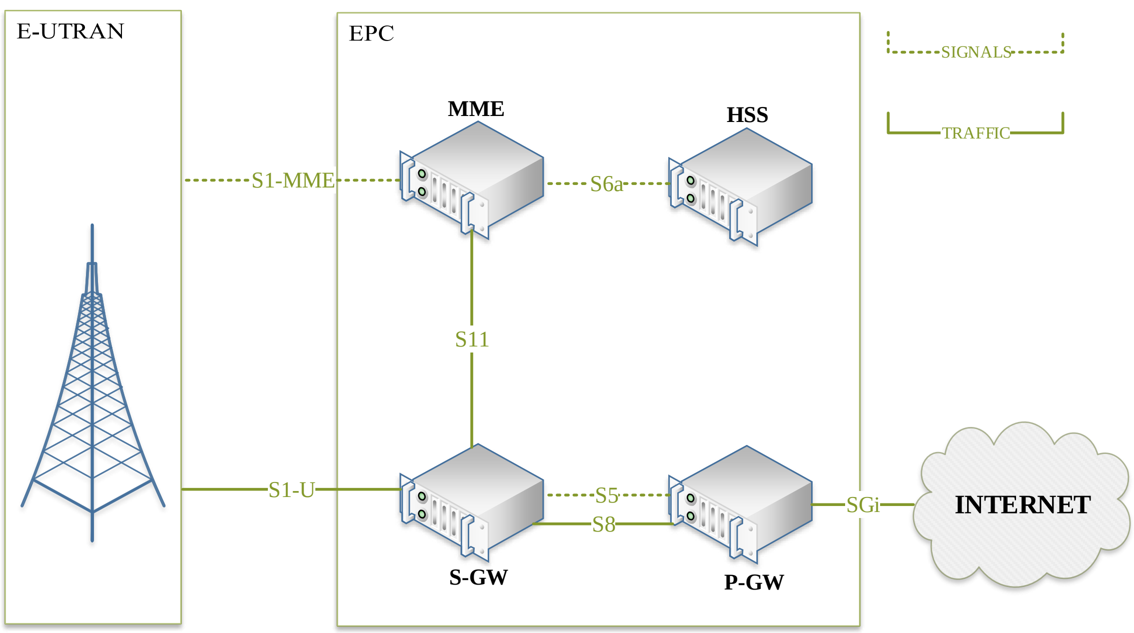


Figure 1.1: Evolved Packet Core (EPC) architecture.

- **Mobility Management Entity (MME)**. The core element of EPC and the main source of intelligence is that in one hand performs a management role of the User

Equipment activities with the core network such as initiating the connectivity to the network, assigning resources, roaming and handovers. Moreover, MME provides various security procedures to protect the user authentication and integrity. On the other hand, it is also compatible with different legacy technologies such as 2G, 3G, and wireless.

- **Policy and Charging Rules Function (PCRF)**. Manages and controls the services dynamically among the users, and applies policy enforcement and flow-based charging via the Policy and Charging Enforcement Function (PCEF).

- **Serving Gateway (S-GW)**. Acts as an access node between eNodeB node and PDN Gateway (P-GW). S-GW also maintains a valid path to the core network to allow mobility of different terminals in the E-UTRN technology or with other 3GPP standards.

- **Packet Data Network (PDN) Gateway (P-GW)**. A gateway that leads to the external data network. In addition, P-GW has PCEF which enforces the operator-defined rules for allocating the resources and filtering the packets.

- **Home Subscriber Server (HSS)**. A centralized data base for storing user information used for authentication and authorization.

## 1.4 Cloud Radio Access Network Architecture

Wireless networks have faced an increasing demand to cope with the exponential growth of data traffic. The conventional architecture of wireless networks has hindered the evolution of network scalability. However, the introduction of cloud technology has brought tremendous flexible and scalable on demand resources. As a result, Cloud Radio Access Networks (C-RANs) have been introduced as a new trend in wireless technologies. C-RAN architecture mainly consists of three parts: Remote Radio Head (RRH), Baseband Unit (BBU), and the optical network that connects them. These essential components are shown in Figure 1.2 and explained in the subsequent subsection.

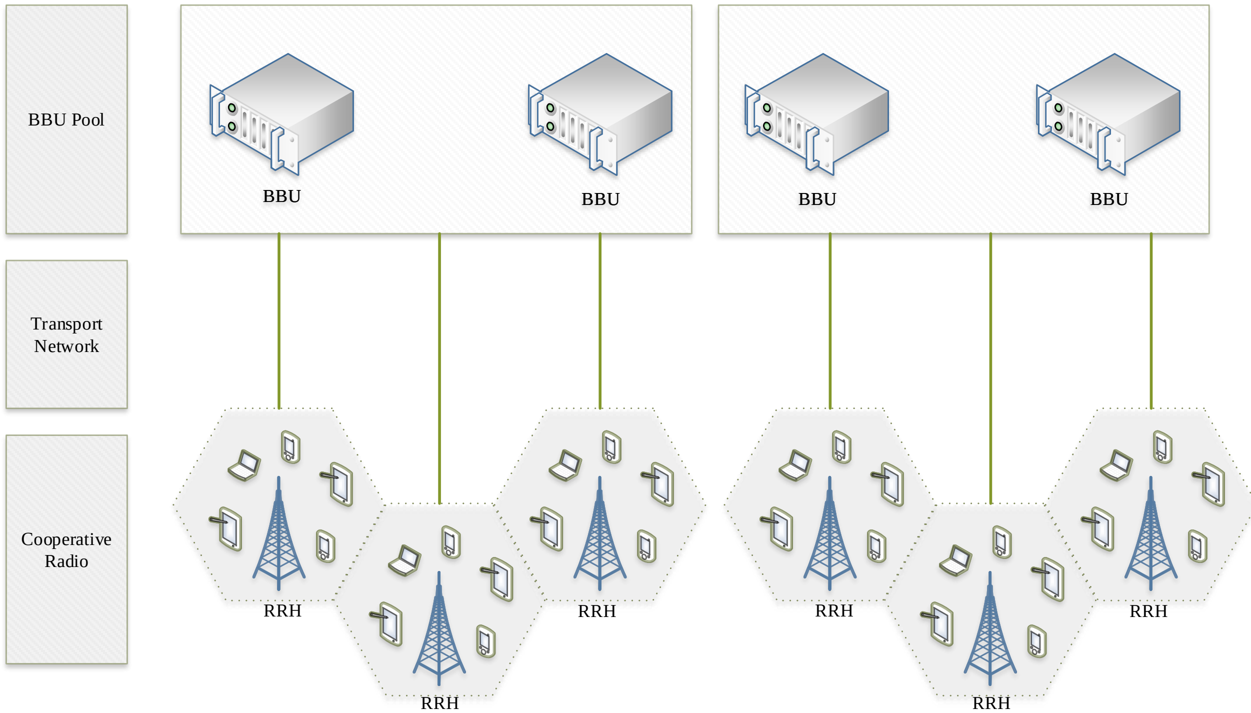


Figure 1.2: Cloud Radio Access Network (C-RAN) architecture.

### 1.4.1 C-RAN Essential Components

C-RAN has on demand centralized processing resource pool that contains multiple processing resources in its own. C-RAN contains three essential components as the following:

1. **Remote Radio Head (RRH)**: RRH is responsible to transmit the radio frequency (RF) signal in the downlink and carries the baseband signal to the virtualized BBU pool in the uplink. RRH has the same role of a simple traditional system of offering the signal coverage for the users in a simplified manner due to the BBU role performing most of the signaling processing operations. Therefore, RRH simple architecture allows having a large number of RRHs distributed with a reduced total CAPEX.

2. **Baseband Unit (BBU) Pool**: Contains set of soft processing BBU nodes each

of which functions to process the basband signals and performs the optimization for the radio resource allocations. The set of the software defined BBUs are located in a centralized site. The processing resources are dynamically allocated based on the scheduling of users and channels' conditions.

3. **Transport Network**: A high-speed transport network that connects between the BBU nodes in the pool and the RRUs. The transport network can be applied through various technologies such as microwave transmission or optical fiber.

## 1.5 Thesis Objectives and Contributions

LTE technology have imported the circuit and packet network of 2G/3G to new flatter IP-network architecture. The new IP paradigm has significantly changed the behavior of control plane signaling. As the base station controller (BSC)/RNC has been eliminated from the new architecture. Instead, the control signaling is distributed between the eNB and MME. This indeed introduces simpler network architecture and more efficient data traffic handling. In spite of all these the advantages mentioned earlier, the signaling has the potential to increase due to the direct connectivity between the E-UTRAN and the MME [15], it is estimated that the average number of messages per subscriber delivered to MME is multiplied by approximately 3 to 4 times when compared to the conventional SGSN [16].

In this context, the signaling overhead is studied for both uplink and downlink. An efficient and novel algorithm that enables a significant reduction of the TAU and paging signaling traffic is defined. To the best of our knowledge, the proposed design is the first that includes MME in the system optimization. Two schemes have been investigated, namely, the distributed MME scheme and the centralized MME scheme. In these schemes, both overlapping TAL assignment and efficient mapping of the MME to the TA/TAL are considered. Furthermore, a heuristic algorithm is presented as a low-complexity approach that gives a sub-optimal solution.

The second part of the thesis intends to enable adaptive online cell-to-TAL assignment in order to further investigate the proposed pooling schemes. UEs are usually in continuous movement and their coordination is not static. Hence, the initial static assignment for cell-to-TAL will gradually become ineffective over a specific period. Therefore, there is a need to revise the TA assignment constantly in order to suit the current mobility state. Unlike conventional TA, the TAL concept allows TA assignment to be modified without interruption of service. This is an advantage of TAL over conventional TA, because TAL provides greater flexibility to the system. Moreover, LTE allows for an auto-reconfiguration feature that adapts the network configuration whenever there is a change in the UE statistics, such as movement patterns and loads.

The third part targets solving the problem of minimizing the signaling overhead using three evolutionary artificial algorithm. The selected algorithms belong to different heuristic search families each of which has been tested for the speed of convergence under different initial random values. Moreover, the reliability of the algorithms are tested based on the relative standard deviation value for the objective function. The three algorithm have proven to be reliable for solving the problem of minimizing the signaling overhead when deployed in a large-scale scenario.

The fourth part discusses C-Radio Access Network infrastructure particularly the Baseband Unit (BBU) where a pool of BBUs gather to process the baseband signals. We argue that there is a need of optimizing the processing capacity in order to minimize the power consumption and increase the overall system efficiency.

A formulated optimization model depicting the two levels of scheduling (from physical resources to users and from RRHs to BBUs) in C-RAN environment is presented. To the best of our knowledge, this study is the first to propose a complete optimization model formulating a strategy for both ends. In the level comprised between cells and users, resources are distributed among users, which have different QoS requirements. As a consequence, the system has to optimize resource allocation accordingly while maintaining other aspects such as availability of physical resources, satisfying the QoS, and continuity of service. In the RRH and BBU level, computing requirements need to be processed instantly in the available BBU pool available while maintaining power consumption and optimizing computing resources. This model is an NP-HARD problem because of the two levels resource assignments. The problem of finding optimal assignment is known to be intractable to be solved in a polynomial time [17, 18]. Therefore, the model is simplified into two sub-problems using the proposed decomposition model. In addition, two heuristic solutions of lower complexity are proposed to solve both problems.

## 1.6 Thesis Organization

The thesis is divided into six chapters. Chapter two introduces a model that relates the tracking area list to the mobility management entity (MME) which enables more control and adds intelligence to the system. Two MME pooling schemes are investigated namely, centralized and distributed MME schemes. The proposed model is NP-hard; thus, the problem can be simplified with a few assumptions (which do not violate the constraints of the problem) to become a solvable linear problem (LP). Moreover, a low-complexity heuristic algorithm is developed by determining the percentage use of the lists/MME in each cell.

Chapter three attempts to improve the intelligence of location management techniques. As an extension of Chapter one, a Self-Organizing Network (SON) that enables dynamic reconfiguration of cell-to-TAL/MME is introduced. A decomposition model that reduces the original formulated problem to two sub-problems is proposed, each of which is solved optimally. In addition, a smart cell-to-TAL selection scheme is proposed to prioritize potential cells that might be visited by a user equipment (UE). Our method is shown to outperform several state-of-the-art methods presented in the literature. Finally, a heuristic algorithm is presented to obtain a less complex solution than the optimal one.

Chapter four focuses on solving the problem of minimization the signaling overhead caused by the location management messages by deploying three evolutionary algorithms namely; particle swarm optimization (PSO), artificial bee colony (ABC), gravitational search algorithm (GSA). Furthermore, the chapter presents a comparison between all of the listed algorithms in terms of objective function minimization and algorithm's speed of convergence when applied to large scale network. The deployed algorithms guarantee yielding the minimum values of the signaling overhead for TAU, paging and the consumed battery power of the user.

Chapter five discusses the optimal allocation of computational resources between the RRHs and BBUs is modeled. Furthermore, in order to get an optimal RRH-to-BBU allocation, it is essential to have an optimal physical resource allocation for users to determine the required computational resources. For this purpose, an optimization problem that models the assignment of resources at these two levels (from physical resources to users and from RRHs to BBUs) is formulated. Due to the high complexity of the formulated problem, a decomposition model was adopted to solve it by formulating two Binary Integer Programming (BIP) sub-problems that model the allocation at each level. Furthermore, two low complexity heuristic algorithms were developed to solve each of the resulting BIP sub-problems.

Chapter six concludes the thesis and highlight a summery of findings.

# Chapter 2
# Intelligent LTE Mobility Management Through MME Pooling

## 2.1 Introduction

The 3rd generation partnership project (3GPP) has developed the long term evolution (LTE) technology to accommodate the increasing data traffic and improve system capacity [19]. LTE offers lower latency, superior speeds, and better quality of service (QoS) performance for mobile networks. A major component of the LTE system is the evolved packet core (EPC) which consists of for main elements: Serving Gateway (S-GW), PDN Gateway (P-GW), Home Subscriber Server (HSS), and the mobility management entity (MME). The MME provides the control-plane support to user equipment (UE) in its domain such as bearer management, paging, handover and tracking area updates [20]. Handling all these functions causes significant signaling overhead at the MME as it needs to handle an average of 290,000 messages per second [6]. This will result an increase of Capital Expenditures (CAPEX) and Operating Expenditures (OPEX) when trying to handle the growing signaling overhead [21]. Two signaling flows dominate the overhead at the MME namely; Paging that is defined as a control message sent by the MME to locate a particular UE in a Tracking Area (TA), and Tracking area update which is another control message sent by UEs to the MME when they move from one tracking area to another.

The concept of tracking area is introduced within LTE. It is defined to be the area in which it is possible to track a UE while it is in the idle mode. A tracking area is formed by grouping a collection of cells together. This concept is known in other technologies. For example, it is called a Routing Area (RA) in GPRS and UMTS and a Location Area (LA) in GSM.

A UE has three basic mobility states:

1. Active mode: the UE can send or receive data to/from the network, and its location is known to the core network.

2. Idle mode: the UE does not send or receive data, and the core network only knows the latest TA that the UE registered to.

3. Detached mode: the UE is in a state of seeking (or registering with) a new cell.

To identify the UE location, the LTE core network pages the latest tracking area that the UE was registered to. The paging signal is received by all of the cells that reside in the same tracking area. Additionally, the UE will update the core network by sending a Tracking Area Update (TAU) signal once it moves from one tracking area to another.

However, there are some limitations that are associated with the conventional TA concept. One inevitable limitation when designing intelligent dynamic cells for TA allocation is the service interruption that occurs whenever a cell is assigned to a new TA, making dynamic allocation impractical to implement [22]. Another drawback is the significant overhead caused by the ping-pong effect of a UE located at the edge between two different TAs. Moreover, in the TA concept, cells cannot have different perspectives toward each other. This means that if cell $A$ considers cell $B$ to be located in the same TA, cell $B$ cannot have a different perception of cell $A$. This considerably limits the flexibility of TA design [22]. Finally, TA has a transitive property. For example, if cell $A$ and $B$ are in the same TA and cell $B$ and $C$ are in the same TA, then cell $C$ and $A$ will explicitly be in the same TA [22].

Release 8 of 3GPP introduced a new concept known as tracking area list (TAL), which groups a set of TAs into lists that are then assigned to UEs. This property mitigates the significant impacts of paging and TAU signaling. Moreover, TALs have a clear advantage over standard TAs in terms of minimizing the ping-pong effect that affects UEs that frequently move from one TA to another. In addition, TALs overcome the TA limitations mentioned above. This introduces a flexible cell to TA allocation and reduces the high uplink traffic. Each cell is responsible for determining the TAL for each UE residing in it. The Mobility Management Entity (MME) is capable of tracking each UE in the latest TAL that the UE was registered to.

In this context, the signaling overhead is studied for both uplink and downlink. An efficient and novel algorithm that enables a significant reduction of the TAU and paging signaling traffic is defined. To the best of our knowledge, the proposed design is the first that includes MME in the system optimization. Two schemes have been investigated, namely, the distributed MME scheme and the centralized MME scheme. In these schemes, both overlapping TAL assignment and efficient mapping of the MME to the TA/TAL are considered. Furthermore, a heuristic algorithm is presented as a low-complexity approach that gives a sub-optimal solution.

It is noteworthy that there is a correlation between TAU and paging: the more cells assigned to a TA/TAL, the smaller the number of TAU storms to the core network (specifically to the MME). In contrast, the paging signaling spreads will increase as the number of cells assigned to the TA/TAL increases.

Several studies have investigated cell-to-TA/TAL assignment problem as they used various techniques to optimize the system by minimizing the signaling overhead. At present, none have investigated TA/TAL-to-MME assignment, which has the same importance as investigating cell-to-TA/TAL assignment. TA/TAL-to-MME allocation is vital for minimizing the signaling overhead gained from the TAU-to-core network, especially in the MME. The MME receives a large amount of signaling traffic resulting from several UEs, each making different requests. As a result, it is more valuable to consider a mechanism to intelligently relate and distribute the signaling load to the MME during cell-to-TA/TAL construction. LTE has introduced the concept of MME pooling, in which the number of clustered MMEs can perform the mobility management function as one MME. Consequently, this decreases the amount of signaling incurred during various procedures such as MME relocation. Ultimately, determining the type of MME pooling versus the TA/TAL allocation will give more accurate control of (and allow optimization of) the signaling overhead resulting from paging and TAU procedures.

This chapter is organized as follows. Section 2.2 introduces the literature works performed within this field. Section 2.3 presents the system model and formulates the problem for both the centralized and distributed schemes. Section 2.4 describes the heuristic algorithm developed to solve the optimization problem. The simulation parameters and results are discussed in Section 2.5. Finally, Section 2.6 concludes the chapter.

## 2.2 Related work

Mobility management in different technologies has been thoroughly investigated in the literature. For instance, a study of GSM technology [23] constructed an analytical model for a location area overlapping mechanism to reduce the signaling cost caused by the ping-pong effect; four selection policies were studied to determine the appropriate percentage of overlapping of LAs. Other works also examined GSM location area mechanisms [24, 25, 26].

In LTE systems, the tracking area list design for large networks has been scarcely explored. In [27], an analysis of LTE mobility management was introduced to evaluate the performance of paging and TAUs in terms of signaling overhead. Three sequential-paging schemes are proposed: paging either the cell itself, the TA of the cell, or the TAL of the cell. Results show that the performance relies on different aspects such as the number of users, the paging overhead, and the polling cycles. In [22, 28, 29, 30], Razavi *et al.* proposed different algorithms for exploiting the TAL concept. An approach for allocating and assigning TALs in large networks is proposed in [22]. The authors built a framework for comparing the TAL concept to the standard TA. The cell load and UE handover statistics were used as metrics to optimize TAL allocation.

In [28], the authors introduced dynamic TAL assignment that takes advantage of LTE's reconfiguration flexibility. The mobile network environment does not always have a steady state in which the load and UE movement remains constant. Frequent changes in the network behavior trigger the need for an adaptable network configuration in which the load of the cells and the TAL assignment are frequently updated. Thus, a dynamic cell-to-TAL assignment that enables optimized assignment for each time interval was suggested. The authors proposed a rule of thumb method that depends on a TAU and paging signaling overhead trade-off. It is claimed that the proposed solution outperforms both conventional TA and the static TAL scheme.

Additionally, a low complexity linear programming model based on overlapping TAL techniques (to mitigate the TAU signaling overhead and to keep the paging overhead under a certain limit) was introduced in [29]. In [30], another overlapping TAL model that linearly optimizes the total signaling overhead at the site level was developed. The study concluded that an optimum solution could be achieved by assigning each site a single TAL that was allocated to each user registered to that site.

In all of the aforementioned studies, the TA has only one cell and most TAL schemes were compared with the conventional TA. In [31], the authors introduced an algorithm to mitigate the signaling burst that is caused by the simultaneous movement of bulk users such as train passengers. The suggested algorithm selectively discovers the location of the TAU burst and creates a rule so as to overcome congestion in the next time interval. In [32], an adaptive algorithm using a movement-based scheme in which the TAL is derived periodically was introduced. The algorithm sets a movement threshold for all TALs and then assigns them to users based on their mobility characteristics.

In [20], a comparison between distributed and centralized MME architectures with respect to the signaling loads was performed. A multicast paging scheme is used instead of the regular unicast paging scheme. Despite the novelty of the idea, the authors did not discuss the formulation of the system in detail. For example, determining the optimal connection of the centralized MME scheme with the TA/TAL formulation is not discussed.

In [33], the authors proposed a solution to support high-speed mobile users by relocating data anchor gateways based on the user mobility and activity pattern. Authors have also introduced handover management mechanism that chooses a suitable target base stations in order to minimize the mobility anchor relocation signaling. In [34], a scheme was presented that contains a set of SGW service areas each of which serves a group of cells or tracking areas. As a result, this can significantly minimize the frequent SGW relocation associated with UE handover.

## 2.3 System model

### 2.3.1 Problem definition

We attempt to determine optimal assignment of overlapping tracking area lists (TALs) for a given set of cells (enB) such that the overall signaling load for both TAU and paging is minimized. Both centralized and distributed MME pool schemes will be examined to determine the performance of both schemes; in the centralized MME scheme, one MME is to be assigned to one TAL , whereas in the distributed MME scheme, one MME is assigned to one TA. However, TA-to-cell assignment is expensive and would cause service interruption in the cell; therefore, it would not be feasible to consider. This section highlights the notation used in this model.

### 2.3.2 Design hypothesis

The following points summarize the design hypothesis that is used in this chapter:

- The number of cells that is assigned to each TA is one. (The terms TA and cell will be used interchangeably.)
- In the distributed MME scheme, the number of TAs that is assigned to the MME is one.
- In the centralized MME scheme, the number of TAs/cells that is assigned to the MME is equal to the number of TA/cells that is assigned to the TAL.
- The tracking update cost is ten times greater than the paging cost, as in [35].
- Cell load are not considered in this design. Figures 2.1 and 2.2 depict the two schemes.

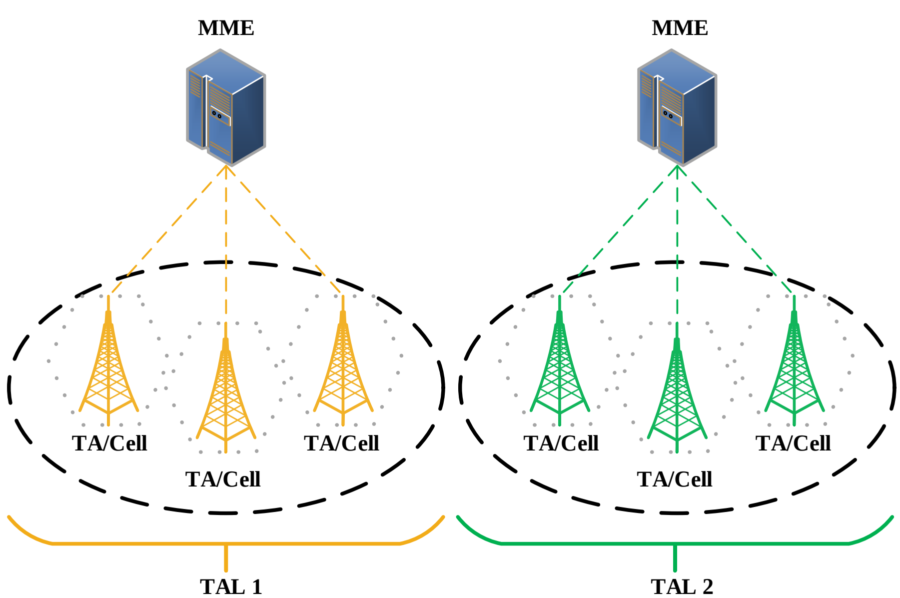


Figure 2.1: Centralized MME pooling scheme

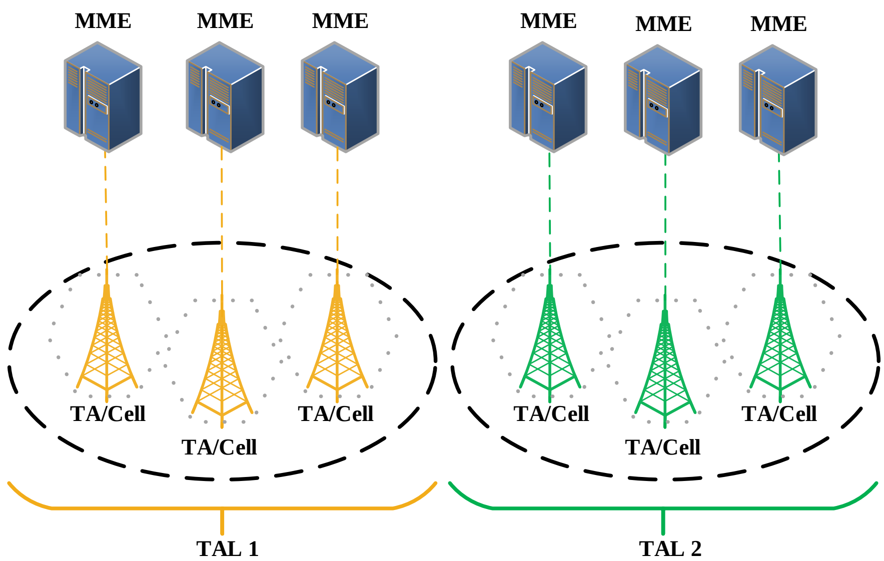


Figure 2.2: Distributed MME pooling scheme

### 2.3.3 Important notation

The following system of notation is used in the system model:

- Decision variables

$$X_{ij}^{lst} = \begin{cases} 1, & \text{if node } i \text{ and } j \text{ belongs to the same list } lst, \\ 0, & \text{Otherwise.} \end{cases}$$

$$O_{lst}^{M} = \begin{cases} 1, & \text{if list } lst \text{ belongs to MME } M, \\ 0, & \text{Otherwise.} \end{cases}$$

$$Y_{i}^{TA} = \begin{cases} 1, & \text{if node } i \text{ belongs to tracking area } TA, \\ 0, & \text{Otherwise.} \end{cases}$$

- Parameters

  $C_{Tu}(i)$: Total signaling cost of TAU in cell $i$.

  $C_u$: TAU cost of UE moving from one list to another.

  $U$ : Average number of UEs.

  $UE_i$: Total number of UEs served by cell $i$.

$\rho$: Paging arrival rate

$H_{ij}$: The probability that a user moves from cell $i$ to cell $j$.

$C_\rho$: Paging cost of particular user equipment.

$C_{T\rho}(i)$: Total paging cost in cell $i$.

$Cu\rho(i)$ : Total paging and TAU overhead in cell $i$.

$\sigma_i^{lst}$: Percentage use of each list $lst$ in cell $i$.

$\omega$: The cost of MME relocation during handover.

$L$: Total number of lists.

$\kappa$: Maximum number of TAs that are assigned to list $lst$.

### 2.3.4 Problem formulation

The problem is addressed using a centralized MME scheme and a distributed MME scheme.

#### 2.3.4.1 Centralized scheme

In this scheme, each TAL is assigned to particular MME. The objective function consists of two parts, as in [30], namely, TAU and paging overhead. In 2.1, the objective function is used to minimize the total amount of signaling overhead for all nodes. The total amount of signaling overhead is calculated by adding the total paging and TAU overhead, as presented in equation 2.2.

$$\min \sum_{i}^{N} Cu\rho(i) \tag{2.1}$$

$$C_{T\rho}(i) + C_{Tu}(i) = Cu\rho(i), \quad \forall i \in N \tag{2.2}$$

Equation 2.3 calculates the total resulting paging signaling overhead in cell $i$ by paging all cells within the same list of cell $i$. The first term calculates the summation of all users in cell $i$ multiplied by the percentile use of each list $lst$ used in node $i$. To clarify the first term further, each user might have different $lst$ depending on the list usage in each cell. The summation checks every list with its usage percentage and multiplies it by the total number of users in that cell. The second term of the equation checks each cell $j$ that is within list $lst$ and multiplies the number of users in that cell by the usage percentage of list $lst$ used in cell $j$. Figure 2.3 depicts an example of the different cells covered by two lists.

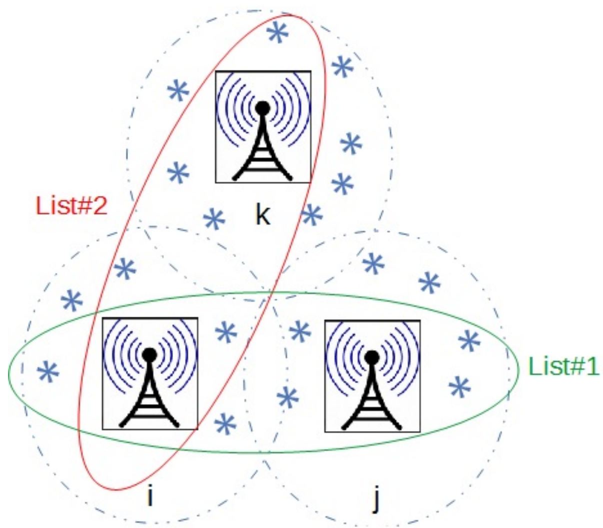


Figure 2.3: Three cells covered by two different lists

$$\begin{aligned} C_{T\rho}(i) = \rho\,.C_\rho(\sum_{lst=1}^{L} UE_i.\sigma_i^{lst} + \\ (\sum_{lst=1}^{L}\sum_{j,i\neq j}^{N} UE_j.\ X_{ij}^{lst}\ .\sigma_j^{lst})), \quad \forall i \in N \end{aligned} \tag{2.3}$$

Equation 2.4 is the total TAU cost function, which is a summation of the probability of users crossing between two different cells that are not in the same list, multiplied by the cost of the tracking update. The value of $\omega$ is added to the cost once the user moves from cell $i$ to $j$.

$$\begin{aligned} C_{Tu}(i) = U.H_{ij}.C_u(\sum_{lst=1}^{L} \omega.\ O_{lst}^{M}.\ \sigma_i^{lst}(1 - X_{ij}^{lst})), \\ \forall i, j, i \neq j \in N \end{aligned} \tag{2.4}$$

Constraint 2.5 ensures that the number of TAs inside the TAL $lst$ do not exceed the limit defined in the standard. Constraint 2.6 ensures fair usage of lists inside each cell $i$, this constraint is vital to ensure a fair usage of MMEs and enhances the system load balance.

In 2.7, the constraint creates a strict policy for any $i$ and $j$ that exist in the list; $j$ and $i$ should also be in the same list. Lastly, 2.8,2.9 and 2.10 are used to create boundaries for the decision variables.

$$\sum_{i}^{N} \sigma_i^{lst}.X_{ij}^{lst} = 1, \quad \forall lst \in L \tag{2.5}$$

$$\sum_{ij,i\neq j}^{N} X_{ij}^{lst} \leqslant \kappa, \quad \forall lst \in L \tag{2.6}$$

$$X_{ij}^{lst} = X_{ji}^{lst}, \quad \forall lst \in L, \ \ \forall i,j,i \neq j \in N \tag{2.7}$$

$$\sigma_i^{lst} \geq 0 \tag{2.8}$$

$$O_{lst}^{M} \in 0,1 \tag{2.9}$$

$$X_{ij}^{lst} \in 0,1 \tag{2.10}$$

#### 2.3.4.2 Distributed scheme

The only difference between the distributed scheme and the centralized scheme is the TAU method, which assigns each TA to specific MME. As a result, an additional signal load will be gained from the MME relocation that is equivalent to the number of TAs multiplied by the signaling load gained from any UEs that move from cell $i$ to $j$, as shown in equation 2.11. In this context, S-GW relocation is not considered.

$$\begin{gathered} C_{Tu} = (U.H_{ij}.C_u(\sum_{lst}^{L} \omega.\ Y_i^{TA} + \\ \sigma_i^{L}(1 - X_{ij}^{lst})) \quad ,\forall i,j,i \neq j \in N \end{gathered} \tag{2.11}$$

The optimization problem described above is a Mixed-Integer Nonlinear Program (MINLP) with quadratic equality constraints. This is a well-known NP-hard problem that cannot be solved to optimality. Therefore, we transform it into a linear program (LP) by making two valid assumptions:

- Assume that each MME handles a single list, and thus, the variable $O_{lst}^{M}$ is known *a priori*.

- Assume that we know the number of TALs and which cells are included in each list, *i.e.*, the values of $X_{ij}^{lst}$, are known *a priori*.

Using these assumptions, we can solve the problem to optimality. Despite the efficient techniques to solve LPs, lower-complexity algorithms can be developed to solve such problems for large scale scenarios.

## 2.4 Heuristic Algorithm

In this section, we develop a heuristic algorithm to find the usage percentage of each list for each cell in the system. In order to justify the need for the heuristic algorithm, we must first discuss the complexity of the decomposition algorithm. The original problem presented is intractable since the model is a mixed-integer nonlinear programming (MINLP) problem with quadratic equality constraints, which is a well-known NP-hard problem that is difficult to solve to optimality. It was assumed that one set of decision variables was known in order to be able to transform the problem to a solvable linear programming problem. This was done for bench-marking purposes. However, such an assumption may not be possible in a realistic environment. Hence, a heuristic algorithm is proposed that solves the problem without making any assumptions about any of the decision variables. The decision variables are determined by finding the number of lists each cell belongs to and sending these lists with equal percentage to the users' equipment residing in the cell. The pseudo-code for the algorithm is depicted in Algorithm 1. The algorithm is divided into 3 phases. In lines 3-10, the number of lists that each cell belongs to is determined by performing an exhaustive search of all lists. The percentage use of each list in each cell is calculated in lines 11-14. The signaling overhead of both schemes is calculated in lines 17-22 using the appropriate equations. In the proposed heuristic solution, the lists that each cell belongs to are used equi-probably. In contrast, the optimal solution distributes each list equally among the cells in it. The complexity of the suggested algorithm is of order $O(LN)$ where $L$ is the number of lists and $N$ is the number of users. The motivation behind the suggested heuristic is that we seek to minimize the signaling overhead caused by tracking area update requests. This is done by sending to each cell all the lists/MME it belongs to, thus increasing the pool of lists/MME, which contain different cells, available for its users. This will reduce the

**Algorithm 1** Heuristic Algorithm

**Input:** $lst = \{1, 2, .., L\}$: all lists
$TA = \{1, 2, ..., N\}$: all tracking areas
$X_{ij}^{lst}$: Cells-to-list binary indicator
**Output:** $\sigma_i^{lst}$: percentage use of list $lst$ in cell $i$
**for** $i \in N$ **do**
  **define** $l_i = \phi$
  **for** $lst \in L$ **do**
    **if** $i \in lst$ **then**
      **update** $l_i = \{l_i \cup lst\}$
    **else**
      **continue**
    **end if**
    **for** $lst \in l_i$ **do**
      $\sigma_i^{lst} = \frac{1}{||l_i||}$
    **end for**
  **end for**
**end for**
**Calculate:**
$C_\rho(i)$ as per equation (2.3)
**if** Centralized Scheme **then**
  **Calculate:**
  $C_{Tu}(i)$ as per equation (2.4)
**else**
  **Calculate:**
  $C_{Tu}(i)$ as per equation (2.11)
**end if**
**Calculate:**
$\sum_i^N Cu\rho(i) = \sum_i^N \big(C_\rho(i) + C_{Tu}(i)\big)$

TAU overhead because it will become less probable that users will send a TAU request while performing handovers.

## 2.5 Performance evaluation

To evaluate the performance of both the centralized scheme and the distributed scheme, we perform MATLAB simulations and compare the results of the optimal and heuristic solutions. The built-in MATLAB function linprog is used to solve the linear programming problem. We study the performance using the signaling overhead of each scheme with respect to two parameters, namely, the number of TALs and the average handover probability. The simulation assumes there are 10 cells, each having a different number of users. The number of users is uniformly distributed with an average of 100. The number of tracking are lists varies between 3 and 5 lists.

Figure 2.4 evaluates the paging overhead performance of both schemes for three different number of lists. We note that both the centralized and distributed schemes have the same paging signaling overhead; this is caused by the cost being the same in both schemes. The same is noted for the heuristic approach. However, the optimal scheme clearly outperforms the heuristic one because the heuristic cares about the number of lists that each cell belongs to rather than the content of each list.

Figure 2.5 shows the TAU cost. It is noted that the centralized scheme outperforms the distributed scheme due to the differing MME relocation cost calculation. In the centralized scheme, this cost is reduced by the percentage use of the list in each cell. However, in the distributed scheme, there is a fixed cost for every MME relocation, thus increasing the signaling overhead. The total signaling overhead is shown in Figure 2.6. It is noticed that it follows the same trend as the TAU. This result is expected because it is assumed that the TAU cost is the dominant cost ($C_{Tu} >> C_{T\rho}$). It is noted that the distributed scheme results are very close to the optimal. This is because of the MME reallocation weight factor has the dominant impact on the equation of the distributed scheme. Thus the gap between the optimal and the heuristic is small. On the other hand, in the centralized scheme, the weight factor is multiplied by the usage percentage of each list resides in the cell. Therefore, the gap is much wider in this scheme.

Figure 2.7 shows the total signaling overhead as a function of the average handover probability. As expected, the signaling overhead increases as the average handover probability increases because the TAU will increase because of the higher number of handovers. Moreover, the centralized scheme outperforms the distributed scheme. As discussed above, this result is due to the fixed cost of the handover. Therefore, the signaling overhead due to the TAU increases as the number of handovers increases, resulting in an increase in total overhead.

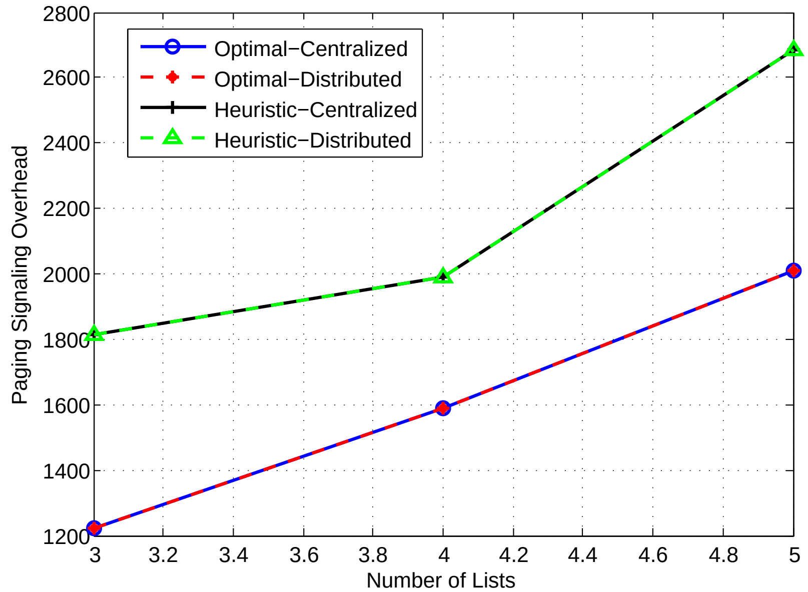


Figure 2.4: Paging signaling overhead cost

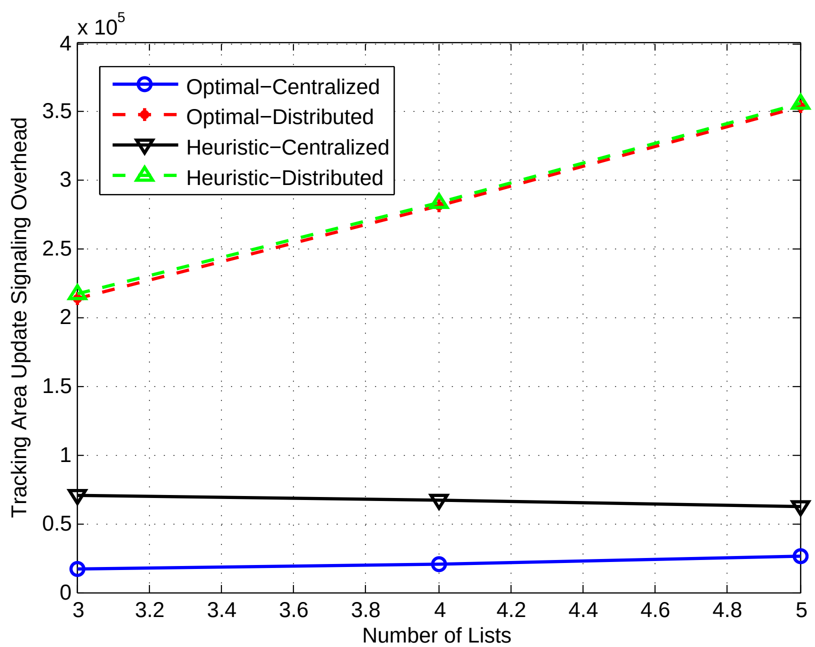


Figure 2.5: TAU signaling overhead cost

## 2.6 Conclusion

Tracking idle users is essential in cellular networks. The concept of the TA is introduced in LTE as the area in which idle UEs can be tracked. The TAL is a new

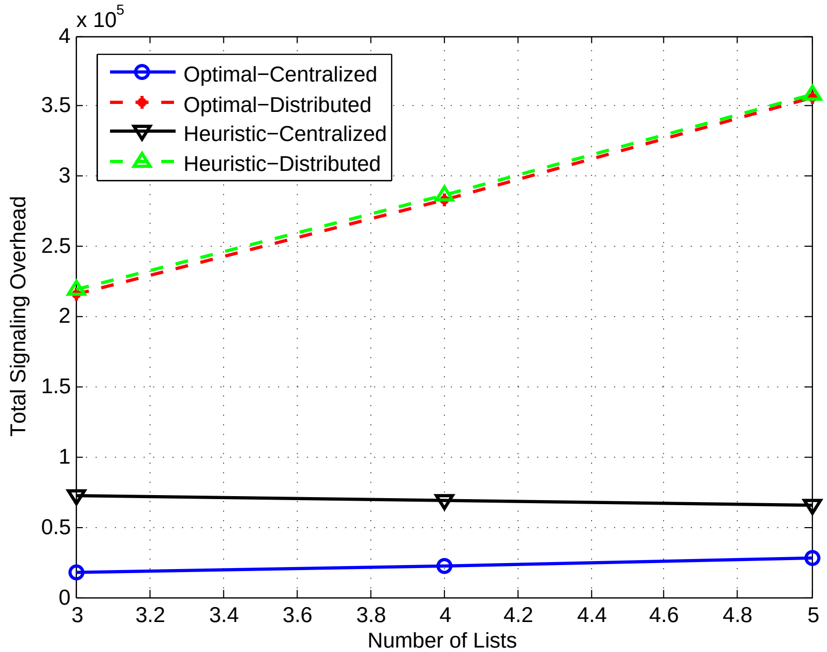


Figure 2.6: Total signaling overhead cost for different number of lists

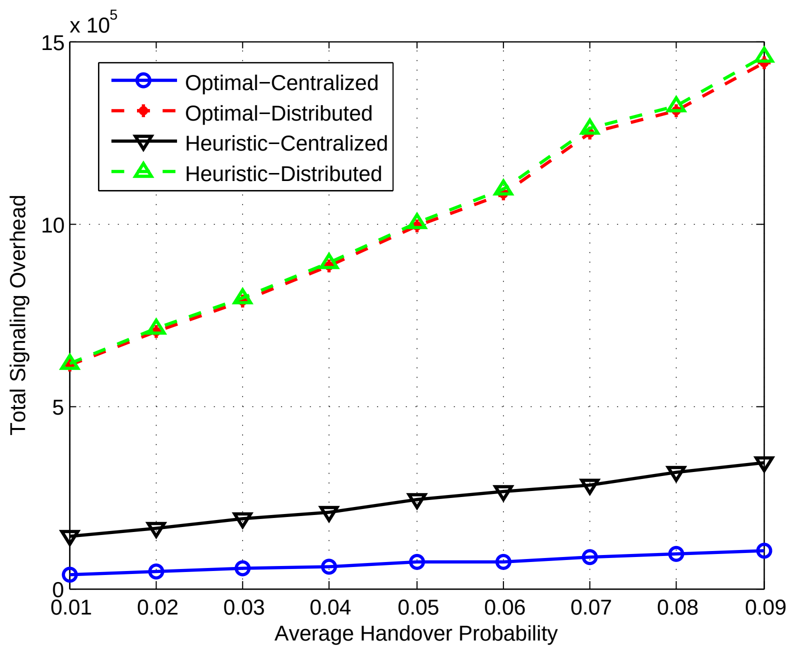


Figure 2.7: Total signaling overhead cost for different handover probabilities

concept introduced in release 8 of 3GPP for LTE networks. A TAL contains a set of TAs that is sent to the UEs in an attempt to reduce both the paging overhead and the

TAU signaling overhead. In this chapter, an optimization problem was formulated that aims to minimize the total signaling overhead. Two schemes were introduced, namely, a centralized scheme that allocates cells to specific lists with each list being handled by a single MME and a distributed scheme that allocates cells to specific lists but with each tracking area being handled by a single MME. The optimization problem is an MINLP that is shown to be NP-hard. Therefore, it is transformed into a linear problem by assuming the knowledge of the cell-to-list allocation. Also, a heuristic algorithm is developed that finds the usage percentage of each list by determining the number of lists that each cell belongs to.

MATLAB simulations showed that the overhead increases as the number of lists increases. Moreover, the centralized scheme outperforms the distributed scheme because the MME relocation cost in the centralized scheme is multiplied by the percentage use of the list in each cell. On the other hand, the relocation cost is fixed in the distributed scheme. Thus, the signaling overhead due to the tracking area update is much larger at the MME level. Furthermore, the optimal solution outperforms the heuristic because the heuristic determines the decision variables by finding the number of lists that each cell belongs to, whereas the optimal solution focuses on the content of each list and distributes it equally among the cells in it.

# Chapter 3
# Dynamic SON-Enabled Location Management in LTE Networks

## 3.1 Introduction

Various mobility management techniques have been investigated extensively from different perspectives, such as overlapping, dynamic, and static cell-to-TAL assignment. The current chapter is an extension of the previous chapter, which introduced cell-to-TAL assignment with two MME pooling schemes. Centralized and distributed MME pooling schemes were investigated in order to explore the difference between both schemes statically [36]. In this context, we intend to enable adaptive online cell-to-TAL assignment in order to further investigate the proposed pooling schemes. UEs are usually in continuous movement and their coordination is not static. Hence, the initial static assignment for cell-to-TAL will gradually become ineffective over a specific period. Therefore, there is a need to revise the TA assignment constantly in order to suit the current mobility state. Unlike conventional TA, the TAL concept allows TA assignment to be modified without interruption of service. This is an advantage of TAL over conventional TA, because TAL provides greater flexibility to the system. Moreover, LTE allows for an auto-reconfiguration feature that adapts the network configuration whenever there is a change in the UE statistics, such as movement patterns and loads.

In release 8, 3GPP introduced the concept of a Self-Organizing Network (SON) that provides a methodology for planning, managing, and optimizing mobile networks in order improve performance efficiency and system reliability. SON has been widely accepted in industry and academia [7, 8]. 3GPP has also released different use cases for LTE, offering self-optimizing and self-healing paradigms. In this context, adaptive TA list management can be used as an SON use case as in [9], which can further reduce the signaling load. In addition, introducing different techniques that relate the TA list to the behavior of the mobile network would further optimize the signaling overhead. Thus, cell-to-TAL assignment can be engineered dynamically while the UE is in continuous movement. The system keeps analyzing the mobility pattern and continuously updates the TA assigned to the list. Thus, the frequency of TAU will be reduced significantly. In this chapter, the mobility pattern is obtained using a fluid flow model to estimate the handover correlation between cells. One way to measure the efficiency of a dynamic system is to examine the UE battery life. Therefore, the UE's battery consumption in static and dynamic techniques will be compared.

The contributions of this chapter can be summarized as follows:

1. A dynamic cell-to-TAL problem is proposed and formulated as a mixed integer non-linear programming (MINLP) problem with quadratic equality constraints. This approach is different from the previous approach presented in [36], where the problem was solved statically by finding the optimal assignment once.

2. The problem is solved using a decomposition model that divides the problem into two sub-problems. The decomposition model allows optimal assignment of cell-to-TAL dynamically instead of having it known a priori in a static fashion as in [36].

3. A new heuristic algorithm differs from the one proposed in the previous chapter and constitutes of two sub-problems in the same manner as the decomposition model. The algorithm dynamically diversifies the TALs among the cells which helps in reducing the TAU signaling load.

4. The proposed dynamic technique is realized through an SON scheme along with a new smart cell selection approach instead of the conventional ring-based cell selection presented in the literature.

5. The proposed algorithms are compared with three state-of-the-art methods presented in the literature : [7],[9], and [37].

The remainder of this chapter is organized as follows. Section 3.2 reviews studies on mobility management techniques. Section 3.3 presents a detailed description of the system model and formulates the problem for both the centralized and the distributed schemes.

Section 3.4 describes the heuristic algorithm developed to solve the optimization problem. Section 5.4 discusses the simulation parameters and results. Finally, Section 3.6 concludes the chapter.

## 3.2 Related Work

A considerable amount of research effort has been devoted to the study of location management in various technologies, specifically in terms of location management, which has not changed much in recent years. In this context, we essentially highlight two major sources of signaling that are directly related to location management: TAU and paging. Numerous studies have discussed the issue of signaling burst caused by TAU and paging. Most of these approaches have investigated the signaling overhead when configuring LA or TA having the same properties. Few researchers have addressed the signaling overhead from the perspective of TAL construction, even though TAL has greater importance and flexibility than LA or TA. In fact, most studies have failed to provide a rigid framework that can provide technical support and is applicable to real-world scenarios.

As an example of previous proposals for older technologies, an overlapping location area mechanism was proposed in [23] for GSM technology. The purpose of the study was to minimize the signaling load resulting from the ping pong effect. The study introduced four selection policies for determining the location area (LA) percentage. Similar related studies have been conducted [24, 25, 26].

Furthermore, TAU and paging techniques have been investigated for the purpose of reducing the signaling load. Several studies [38, 39, 40] have discussed a number of methods, such as timer-, velocity-, and movement-based ones. However, the aforementioned methods are not commonly used in the current TA/TAL approach.

Few studies have discussed the signaling load in terms of steering the load through the control-plane elements, such as MME or service gateway (S-GW). In one study [20], a concept that enables two modes of MME distribution, namely centralized and distributed MME, has been introduced. The study proposed analysis of both architectures in terms of the signaling load resulting from user mobility. The authors also presented a comparison between multicast and unicast paging. While the study reported some findings using distributed and centralized MME architectures, it did not discuss the fundamentals of

TA/TAL construction that allow further exploration of both architectures. Furthermore, the authors did not illustrate how the MMEs are allocated and implemented in their model. Another study [33] has proposed a model that supports high-mobility users by relocating data plane gateways on the basis of the UE's mobility pattern, thereby minimizing the relocation frequency.

In [34], the authors defined the concept of S-GW service area, where a pool of S-GWs serves TAs or cells. The objective is to eliminate frequent disconnections that occur when the UE moves to a different cell or TA. The concept focuses specifically on active users whose Quality of Service (QoS) can be degraded significantly during S-GW re-allocation.

A number of studies have proposed dynamic TA/TAL techniques that can optimize the signaling overhead periodically [28, 32, 41]. In [28], the authors explored the advantage of dynamic TAL configuration by introducing a "rule of thumb". The approach attempts to minimize the signaling overhead on the basis of the TAU and paging correlation. It succeeded in reducing the overhead to a greater extent than conventional TA. In the proposed model, dynamic configuration is achieved after a fixed period, which does not yield accurate results in environments having UEs moving at different speeds. In [32], the authors introduced an adaptive model that constructs a suitable TAL for each user in a set of cells represented as an individual TA. The adaptive configuration is triggered periodically on the basis of a defined movement threshold. This model does not clearly indicate the feasibility of the solution when dealing with a large-scale scenario. Moreover, the authors did not provide a detailed description of the physical parameters of the system, such as the cell radius and the users' velocity. In [41], the authors presented a dynamic model that allows TAs to be configured as user mobility patterns change. The algorithm tends to use the graph correlation coefficient method, which measures the similarities in user mobility behavior within a certain period and triggers new TA construction. The approach is expensive owing to the high cell-to-TA reconfiguration cost.

Other studies have deployed TA/TAL configuration as an SON use case, e.g., [9]. The system implements procedures and protocols based on the UE behavior pattern. Consequently, the TA/TAL is configured periodically to minimize the signaling burst. The study proposed adapting the TAL configuration depending on the user mobility pattern using a set of femto-cell mesh networks. Each TAL can be configured dynamically on the basis of the UE's mobility status. Despite the novelty of the algorithm, the model does not specify some major design aspects, such as the advantage of overlapping TALs in the cells, which can contribute significantly toward minimizing the signaling load.

TAL has been explored and modeled in [22, 29, 30, 27, 7]. In [22], the authors proposed a model for TAL that can be applied to a large-scale scenario, and they showed the advantages of TAL over the conventional TA concept. Thus, they tried to provide an abstraction of the TAL design and its benefits. In [29], the same authors proposed an optimization model based on overlapping TALs, which would allow a cell to distribute different TAL portions to a set of users residing in that cell in order to alleviate the TAU signaling load whilst restricting the paging load within a defined limit. In [30], the authors formulated a linear programming (LP) model based on overlapping TALs in large-scale scenarios and compared it with the conventional TA technique. Mobility management in LTE was also investigated in [27]. Paging and TAU were analyzed in terms of signaling. Three sequential paging schemes were presented, namely cell-TAL, TA-TAL, and cell-TA-TAL, whereby the MME first requests the cell to page the UE, and if the cell fails to allocate the UE, the MME sends another request to all cells residing in the TA or TAL. This can also be done in a different order. The authors concluded that the results vary with the number of users, paging schemes, and polling cycles. Thus, the selection of potential paging schemes will depend on several factors, such as minimizing the signaling cost of paging and TAU or reducing the number of polling cycles. Lastly, the performance of TAL was modeled and analyzed in [7]. The study investigated TAL modeling from several aspects, e.g., limited number of TAs in the TAL, and introduced two call-handling models that can provide a more realistic view of the system. The total signaling cost was calculated and the optimal TAL for UEs was determined. However, the authors did not verify the validity of the model in large-scale scenarios. Moreover, the model did not emphasize the design of overlapping TALs or the distribution of TALs within the system.

## 3.3 System Model

### 3.3.1 Problem Definition

The objective of the problem is to minimize the signaling overhead due to TAU and paging in the core network. Furthermore, our model seeks to achieve load balancing through different MMEs. The same concept presented in our previous chapter [36] will be applied along with an SON dynamic algorithm; the latter enables periodic cell-to-TAL/MME configuration. The model consists of two pooling schemes, namely centralized and distributed schemes. The following table (Table 3.1) summarizes the notations used in the system model.

### 3.3.2 Preliminaries

In order to design a valid model for minimizing the signaling load, we first show that paging and TAU have a tight correlation that can be used for constructing a TAL. A TAL consists of a number of TAs, each of which can be represented as one cell. The size of each TAL has a direct influence in terms of signaling. In other words, as the number of cells accommodated inside a TAL increases, the paging signaling load increases and the TAU signaling load decreases, and vice versa. This is a result of the TAU and paging mechanisms. A TAU signal is triggered to update the user's location within the MME. This is done whenever a user moves to a cell that is in a different TAL. A paging signal is triggered to locate the user by messaging all the cells of the last TAL to which the user was registered. As the number of cells within a list increases, fewer TAU are required because the probability that the user changes the list has decreased. However, paging would increase because the MME would need to send messages to a larger group of cells in order to accurately determine the location of the user. Thus, the signaling overheads due to TAU and paging are inversely proportional. This can be clarified in the following equations 3.1 and 3.2. Let us assume that $X_{ij}^{l}$ is the cell-to-list assignment, where cells $i$ and $j$ reside in list $l$. Further, the constant costs of paging and TAU are denoted by $C_\rho$ and $C_u$, respectively. We can now express the initial cost functions of paging and TAU caused by a UE as follows:

Table 3.1: Table of Notations

| | | |
|---|---|---|
| $C_{Tu}(i)$ | : | Total signaling cost of TAU in cell $i$. |
| $C_u$ | : | TAU cost of UE moving from one list to another. |
| $UE_i$ | : | Total number of UEs served by cell $i$. |
| $\rho$ | : | Paging arrival rate. |
| $U$ | : | Average number of UEs. |
| $H_{ij}$ | : | Probability that a user moves from cell $i$ to cell $j$. |
| $C_\rho$ | : | Paging cost of particular user equipment. |
| $C_{T\rho}(i)$ | : | Total paging cost in cell $i$. |
| $Cu\rho(i)$ | : | Total paging and TAU overhead in cell $i$. |
| $L$ | : | Total number of lists. |
| $N$ | : | Total number of cells. |
| $lst$ | : | Individual list. |
| $\kappa$ | : | Maximum number of TAs that are assigned to list $lst$. |
| $\omega$ | : | Cost of MME relocation during handover. |
| $HX^{lst}(i)$ | : | Inter-list handover rate of users in cell $i$. |
| $O_{lst}^{M}$ | = | $\begin{cases} 1, & \text{if list } lst \text{ belongs to MME } M, \\ 0, & \text{otherwise} \end{cases}$ |
| $Y_i^{TA}$ | = | $\begin{cases} 1, & \text{if node } i \text{ belongs to } TA, \\ 0, & \text{otherwise} \end{cases}$ |
| | | Decision Variables |
| $\sigma_i^{lst}$ | : | Usage ratio of each list $lst$ in cell $i$. |
| $X_{ij}^{lst}$ | = | $\begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ belong to } lst, \\ 0, & \text{otherwise} \end{cases}$ |

- The cost function of paging a specific UE within a certain cell $i$ is given by

$$C_\rho * \sum_{l}^{L}\left(\sigma_i^l + \sum_{j=1,j\neq i}^{N} \sigma_j^l X_{ij}^l\right) \tag{3.1}$$

  In order to clarify the equation above (3.1), let us assume that cell 1 and 2 are in the same list 1, the value of $X_{12}^1$ is set to 1, all other variables are either set to one or zero depending on whether or not they belong to the same list. Equation (3.1) calculates the cost function of paging a specific user belonging to particular cell which is cell 1 in this example. Thus, variable $X_{12}^1$ would be multiplied by the usage ratio of list 1 in cell 1 (denoted as $\sigma_1^1$) as well as a constant value of the paging cost $C_\rho$. The value of $\sigma_1^1$ is 1 since cell 1 only belongs to one list which is the only list provided to the cell. Similarly, $\sigma_2^1$ is also one since cell 2 belongs to one list only. Therefore, the core network would page every cell belonging to list 1 that the intended user is within (cells 1 and 2 in this example).

- The cost function of TAU generated by a specific user crossing two cells that are not within the same TALs $l$ is given by

$$C_u * \sum_{l}^{L} \sigma_i^l (1 - X_{ij}^l) \tag{3.2}$$

  Equation (3.2) is to calculate the tracking area update cost generated by a specific user moving from cell $i$ to cell $j$. Let us assume that a user moves from cell 2 to cell 3 that is not within the same list (e.g, cell 2 belongs to list 1 and cell 3 belongs to list 2). The TAU cost would be multiplied by one for each two cells that are not within the same list. Note that the same reasoning given for the paging cost is used here for the value of $\sigma$ and hence, $\sigma_2^1$ is set to 1 whereas $\sigma_2^2$ is set to 0 because cell 2 belongs to list 1 and not list 2. Therefore, the user would incur a signaling overhead equal to the TAU cost $C_u$.

### 3.3.3 Design Hypothesis

As stated previously, the model consists of two MME pooling schemes. The centralized MME pooling scheme allocates each TAL to an individual MME element. On the other hand, the distributed MME pooling scheme allocates each TA or cell to an individual MME element. Note that in our architecture, a TA always contains only one cell; this simplifies the problem and avoids the cell-to-TA assignment constraint. Further, in this context, a cell or TA can be used interchangeably and will have the same constraint.

The model have two levels of configuration: cell/TA-to-TAL/MME assignment, which can be handled periodically by the core network as illustrated in the first sub-problem, and TAL/MME-to-UE assignment, which will be handled through each cell as explained in the second sub-problem. First, in terms of cell-to-TAL/MME assignment, the system considers the mobility pattern of the UEs within the cells and dynamically assigns or re-assigns the cells inside the TALs in order to minimize the TAU signaling overhead. Second, in terms of TAL-to-UE assignment, each node will assign a number of TALs to a portion of UEs residing within that cell with a defined usage ratio, denoted by $\sigma_i^{lst}$. The decision variable $\sigma_i^{lst}$ determines the usage ratio of each TAL within each cell. This results in overlapping TAL assignment to cells. Furthermore, this technique can also provide load balancing through the MMEs in the centralized pooling scheme, because each MME represents an individual TAL.

#### 3.3.3.1 Cell/TA-to-TAL/MME Assignment

We attempt to determine the optimal cell/TA-to-TAL/MME assignment in the centralized pooling scheme, while the distributed scheme considers cell-to-TAL assignment only. In the centralized scheme, each TAL can be allocated to each MME on a one-to-one basis. Alternatively, in the distributed scheme, each cell is represented as a TA that is assigned to an individual MME. LTE has defined the maximum number of TAs that can be allocated to a TAL to be 16. As a basic definition for the assignment constraint, we can use the binary decision variable $X_{ij}^{lst}$ to determine whether cells $i$ and $j$ reside in the same tracking area list. The constant $\kappa$ is the maximum number of TAs in the TAL as described in the following equation:

$$\sum_{i}^{N} \sum_{j,j\neq i}^{N} X_{ij}^{lst} \leqslant \kappa, \quad \forall lst \in L \tag{3.3}$$

The purpose of the above-mentioned constraint is to allocate a maximum of 16 cells/TAs to each TAL. However, to eliminate the redundancy of cell/TA combinations in each TAL, another constraint should be considered as follows:

$$X_{ij}^{lst} = X_{ji}^{lst}, \quad \forall lst \in L, \ \ \forall i, j, i \neq j \in N \tag{3.4}$$

We can represent $X_{ij}^{lst}$ as a matrix having different cell/TA-to-TAL/MME combinations. To describe the search space that $X_{ij}^{lst}$ can have, let us assume that the maximum number of cells/TAs in a list is 3 and that there are 3 lists or MMEs. Therefore, there will be 3 x 3 possible combinations of the binary variable $X_{ij}^{lst}$ generated in the system. The following matrix shows the choices that list $lst$ can have.

$$X_{ij}^{lst} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \tag{3.5}$$

From the aforementioned information, the number of cell-to-TAL combinations can be generalized as $L \times N$. In this context, a dynamic algorithm is proposed to periodically update the cell-to-TAL configuration. A cell can have a number of neighboring cells that are initially arranged as rings. Numerous studies, e.g., [7],[9],[37] have considered periodically expanding or eliminating the cells arranged as rings. In fact, the ring-based assumption is not entirely efficient in terms of minimizing the signaling load. This is because the concept of rings might involve cells that have a low probability of being visited owing to their mobility behavior. Conversely, our technique involves analyzing the average mobility behavior of the UEs over a certain period and expanding or eliminating specific cells not strictly included in the entire ring. In other words, the proposed technique selects a cell to be included or eliminated from a pool of cells that is not restricted directly to the surrounding neighbors of the current cell. This can significantly minimize the signaling load caused by frequent TAL updates and lead to the inclusion of a greater number of potential cells that are located outside the ring and have a higher probability of being visited by the UEs. Thus, it is an efficient scheme whereby cells can be chosen selectively rather than in terms of group of rings. Figure 3.1 shows the difference between the proposed smart selection method and the ring selection method. We note that the smart technique allows different selection shapes, such as the one depicted in orange. On the other hand, the ring selection technique, shown in green, can only include all the surrounding cells, regardless of their probability of being visited.

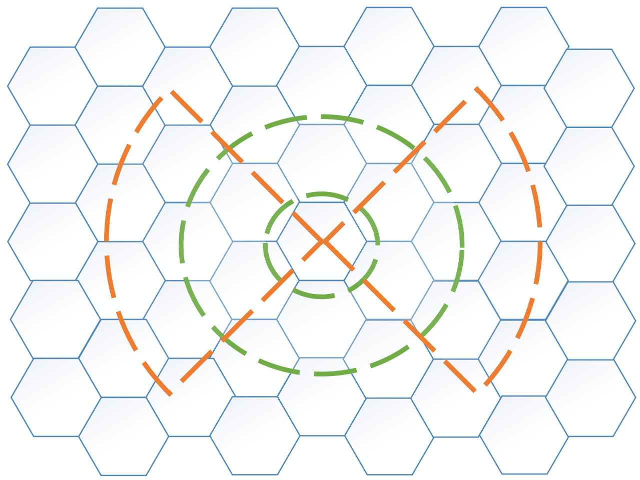

Figure 3.1: Ring neighbor selection versus smart cell selection

#### 3.3.3.2 TAL-to-UE

Each TAL is assigned to UEs via their serving cells. TALs are distributed among the users based on the number of lists that the cell belongs to. This is because if a cell is assigned to multiple lists, the UEs need to have access to these multiple lists.

### 3.3.4 Problem Formulation

The original model that was proposed in previous chapter [36] is used as a base model for both the centralized and the distributed pooling schemes. The objective of the model is to minimize the signaling overhead resulting from both TAU and paging. The model is modified and divided into a bi-objective minimization problem as follows:

- **Objective function**

$$\min \quad \alpha \sum_{i}^{N} Cu\rho(i) + \beta \sum_{i}^{N} HX^{lst}(i) \tag{3.6a}$$

- **Cost functions**

$$Cu\rho(i) = C_{T\rho}(i) + C_{Tu}(i), \quad \forall\, i \in N \tag{3.6b}$$

$$HX^{lst}(i) = \sum_{lst=1}^{L} UE_i.H_{i,j}.(1 - X_{ij}^{lst}) \tag{3.6c}$$

- **Centralized pooling-TAU cost**

$$\begin{aligned} C_{Tu}(i) = UE_i.H_{ij}.C_u(\sum_{lst=1}^{L} \omega.\ O_{lst}^{M}.\ \sigma_i^{lst}(1 - X_{ij}^{lst})), \\ \forall\, i, j, i \neq j \in N \end{aligned} \tag{3.6d}$$

- **Distributed pooling-TAU cost**

$$\begin{aligned} C_{Tu}(i) = (UE_i.H_{ij}.C_u(\sum_{lst}^{L} \omega.\ Y_i^{TA} + \\ \sigma_i^{L}(1 - X_{ij}^{lst})) \quad , \forall\, i, j, i \neq j \in N \end{aligned} \tag{3.6e}$$

- **Paging cost**

$$\begin{aligned} C_{T\rho}(i) = \rho\ .C_\rho(\sum_{lst=1}^{L} UE_i.\sigma_i^{lst} + \\ (\sum_{lst=1}^{L} \sum_{j,i\neq j}^{N} UE_j.\ X_{ij}^{lst}\ .\sigma_j^{lst})), \quad \forall\, i \in N \end{aligned} \tag{3.6f}$$

- **Constraints**

$$\sum_{lst=1}^{L} X_{ij}^{lst}.\sigma_i^{l} = 1, \quad \forall\, i \in N \tag{3.6g}$$

$$\sum_{ij,i\neq j}^{N} X_{ij}^{lst} \leqslant \kappa, \quad \forall\, lst \in L \tag{3.6h}$$

$$X_{ij}^{lst} = X_{ji}^{lst}, \quad \forall lst \in L, \ \ \forall\, i, j, i \neq j \in N \tag{3.6i}$$

$$0 \leq \sigma_i^{lst} \leq 1 \tag{3.6j}$$

$$O_{lst}^{M} \in 0, 1 \tag{3.6k}$$

$$X_{ij}^{lst} \in 0, 1 \tag{3.6l}$$

The objective function (3.6a) is divided into two parts. The first part minimizes the signaling overhead caused by TAU and paging and the second part minimizes the inter-list handover. Both parts have equivalent weights in terms of importance and are represented as $\alpha$ and $\beta$. The first part of (3.6a) contains two cost functions, as shown in (3.6b), which combines the cost function of TAU in the centralized scheme, as in (3.6d), and that in the distributed scheme, as in (3.6e), along with the second term, which has the paging cost function, as in (3.6f). In the centralized scheme, the signaling load is determined by first calculating the number of UEs that reside in a cell $i$ and have a probability of moving to another cell that is not within the same list, and then multiplying that number by the cost of inter-MME reallocation ($\omega$) and the usage ratio of each list ($\sigma_i^{lst}$). In the distributed scheme, the cost of MME reallocation is multiplied by the number of UEs that travel from one tracking area to another. The paging cost function (3.6f) calculates the signaling load of paging messages that are triggered by the UEs inside the lists. The cost considers the percentage of the overlapping lists used in cell $i$ multiplied by the decision variable $X_{ij}^{lst}$ that determines whether cell $i$ and the neighboring cell $j$ belong to the same list. The second part of (3.6a) considers estimating the inter-list handover rate of the average number of UEs in a cell $i$, as given in (3.6c). Constraint (3.6g) ensures fair usage by a set of cells for every list/MME. Moreover, the load of MMEs is balanced throughout the cells under each assigned MME. Constraints (3.6j), (3.6k), and (3.6l) are the boundary constraints.

### 3.3.5 Decomposition Model

The previous model is a mixed-integer nonlinear programming (MINLP) problem with quadratic equality constraints. This problem is a well-known NP-hard problem that is difficult to solve to optimality [42]. In our previous chapter, the binary decision variable was defined in the model and cell-to-TAL/MME assignment was done statically.

In this context, we want to have dynamic cell-to-TAL/MME assignment based on the UE mobility patterns. Therefore, we propose a decomposition model that consists of two sub-problems, each of which is defined as follows:

**Sub-problem 1**. *In a given set of cells that serve a number of UEs, find the optimum cell-to-TAL/MME assignment allocated periodically to minimize the inter-list handover rate of UEs traveling from a cell to another cell that is not within the same list TAL.*

**Sub-problem 2**. *In a given set of TALs/MMEs that are overlapping and used by a number of cells, find the optimum usage ratio of each TAL/MME to be given to a number of UEs that are located inside a cell such that the signaling overhead is minimized.*

#### 3.3.5.1 Sub-problem 1 Formulation

$$\min \sum_{i}^{N} HX^{lst}(i) \tag{3.7a}$$

$$HX^{lst}(i) = \sum_{lst=1}^{L} UE_i.H_{i,j}.(1 - X_{ij}^{lst}), \;\; \forall\, i, j, i \neq j \in N \tag{3.7b}$$

$$\sum_{i}^{N} X_{ij}^{lst} \geq 1, \;\; \forall\, lst \in L \tag{3.7c}$$

$$(3.6h) \text{ and } (3.6i) \tag{3.7d}$$

$$(H_{i,j} - Limit).X_{ij}^{lst} \geq 0, \; \forall\, i, j, i \neq j \in N, \; \forall\, lst \in L \tag{3.7e}$$

The objective function (3.7a) tries to create a suitable $X_{ij}^{lst}$ that tends to minimize the occurrence of UEs moving between cells that are not within the same TAL. The cost function (3.7b) calculates the cost of the handover rate of UEs crossing cells belonging to different TALs. Constraint (3.7c) assigns at least two cells to every list $lst$. Constraint (3.7d) is used to satisfy the conditions of $X_{ij}^{lst}$ formulation and limit the number of TAs inside each TAL. Finally, constraint (3.7e) prioritizes the cells having a higher probability of being visited for inclusion in the same list, as indicated by the UE mobility patterns. This constraint is calibrated and modified on the basis of the UE average speeds.

#### 3.3.5.2 Sub-problem 2 Formulation

$$\min \sum_{i}^{N} Cu\rho(i) \tag{3.8a}$$

$$Cu\rho(i) = C_{T\rho}(i) + C_{Tu}(i), \ \ \forall\, i \in N \tag{3.8b}$$

$$C_{Tu}(i) \ = \ (3.6d) \text{ or } (3.6e) \text{ based on pooling scheme} \tag{3.8c}$$

$$\begin{aligned} C_{T\rho}(i) = \rho\ .C_\rho( \sum_{lst=1}^{L} UE_i.\sigma_i^{lst} + \\ (\sum_{lst=1}^{L} \sum_{j,i\neq j}^{N} UE_j.\ X_{ij}^{lst}\ .\sigma_j^{lst})), \ \ \forall\, i \in N \end{aligned} \tag{3.8d}$$

$$\sum_{lst=1}^{L} X_{ij}^{lst}.\sigma_i^{l} = 1, \quad \forall\, i \in N \tag{3.8e}$$

$$0 \leq \sigma_i^{lst} \leq 1 \tag{3.8f}$$

$$X_{ij}^{lst} \in 0, 1 \tag{3.8g}$$

The description of the objective function and significance of the constraints have been addressed in our previous subsection.

### 3.3.6 Mobility Pattern Model

We use the fluid flow model to simulate the mobility behavior of the users in the system. The fluid flow model is a well-known model that is commonly used in the literature. The model depicts the traffic flow rates of UEs moving out of a closed region represented as a cell or base station. For a certain cell $i$ with perimeter $L$, UE density $Ui$, and average UE velocity $v$, the average number of cell crossings per unit time is calculated as follows:

$$\frac{Ui.L.v}{\pi} \tag{3.9}$$

In this context, cells are hexagonal in shape with side length $R$; hence, $L$ is replaced with $6R$.

### 3.3.7 SON Capability through MME

Self-Organizing Network is a paradigm that seeks to minimize the operation expenses related to network re-organization in order to achieve higher efficiency and Quality of Experience (QoE) [43]. However, it is vital to deliver an updated version of TAL due to the rapid changes in the mobility patterns of the UEs. The MME can be involved in triggering the self-optimizing capability through the network by sending the updated TALs to the base stations. The purpose of dynamic cell-to-TAL/MME reassignment is to alleviate the signaling overhead that results specifically from the movement of UEs from one cell to another cell that is not in the same TAL/MME. The dynamic algorithm will set a timer to trigger the desired changes in the TALs accurately and to distribute them among the cells. Two vital factors tend to have a major impact on the algorithm: the average handover rate of the UEs between the cells, and the given value of LIMIT that prioritizes the combinations of cells to be allocated within the same list or MME. The parameter LIMIT can be chosen by the network operator as a control parameter to prioritize cells over others in order to reduce the required TAU. Moreover, the value of LIMIT is related to the average handover rate between cells or the average UE velocity. Equation (3.7e) shows the relation between LIMIT and the average handover rate. We can see that as the value of LIMIT increases, a greater number of cells having a higher handover rate between each other are included in the same list. In this context, we assumed that LIMIT can have different values related to the UE average velocity. A detailed description of the dynamic algorithm is provided in the next subsection.

### 3.3.8 Dynamic TAL Algorithm

The proposed dynamic algorithm mainly focuses on Sub-problem 1, which plays the role of allocating suitable combinations of cells in each TAL/MME. This tends to minimize the handover rate between cells that are not in the same TAL/MME. As mentioned earlier, a timer will be used to trigger the frequent update of cell-to-TAL/MME assignment. Furthermore, the TAL/MME is updated within a selective set of cells that are determined by the value of LIMIT, which varies periodically corresponding to the UE mobility patterns. The timer is adjusted by $\frac{R}{v}$, which determines the average time required for the UEs to cross a certain cell. This defines the problem's time horizon as the optimization problem takes place within the system. Pseudo-code depicted in Algorithm 2 explains the mechanism of the proposed SON dynamic algorithm in the model. In line 1, the algorithm initializes the input values, which include the number of TALs/MMEs used in the system and the number of TAs, which is equivalent to the number of cells. Further, both the average velocity rate and the average number of UEs in each cell are determined. Finally, the cell radius and the initial cell-to-list assignment are given. The output of this algorithm is the optimized cell-to-TAL/MME assignment. In lines 4-6, at a given time, the algorithm calculates the average crossing rate of UEs residing in each cell in order to determine the handover rate $H_{ij}$ between the cells. Then, the LIMIT value is calibrated to selectively choose the neighboring cells of any cell $i$ that have the highest handover rate and allocate those neighboring cells to the same list or MME. In lines 7-8, the initial assignment is modified by either eliminating or retaining the old cells that had previously been allocated to the lists. This is achieved by solving Sub-problem 1. In lines 9-15, a timer is set to estimate the average time needed for a UE to leave the cells. The timer determines the optimal frequency with which the model should reassign the cells to the lists. Moreover, the system randomly changes the values of the UEs in every event prior to the trigger time and keeps calculating the signaling cost by solving Sub-problem 2.

## 3.4 Heuristic Algorithm

### 3.4.1 Algorithm Description

This section describes the development of a heuristic algorithm that depends on equal distribution of the TAL/MME load among the relevant cells. This approach has the advantage of minimizing the TAUs among the cells in a less complex manner. In this context, the current heuristic algorithm differs from the preceding one presented in [36]; it follows the same concept as the optimal formulation, which divides the original algorithm into two sub-problems. The first sub-problem provides a new method for assigning cells to TALs/MMEs by selecting the cells having the highest values of UE crossing rates such that the TAU frequency is minimized. The cell-to-list assignment will also be performed dynamically by defining the timer in the manner described above for

**Algorithm 2** SON Dynamic Algorithm

**Input:** $lst = \{1, 2, ..., L\}$: All lists = MMEs
    $TA = \{1, 2, ..., N\}$: All tracking areas
    $UE$: Average number of UEs in cell $i$
    $v_{Range} = [v_0 : v_{max}]$
    $v = Random$ such that $v \in v_{Range}$
    $Ui$: Density of UEs in cell $i$
    $R$: Cell diameter
    $UE_i$: Initial number of UEs in cell $i \in TA$.
    $X_{ij}^{lst}$: Initial random cell-to-TAL/MME assignment
**Output:** $X_{ij}^{lst}$: Optimized cell-to-TAL/MME assignment
**for** $t \in TotalTime$ **do**
    **define** $UE_r^{ij} = \frac{Ui.L.v}{\Pi}$: Average crossing from $i$ to $j$
    **define** $Hij = \frac{UE_r^{ij}}{UEi}$
    **update** $LIMIT$
    **Calculate Sub-problem 1:**
    **update** $X_{ij}^{lst}$
    **for** $t' \in (R/v)$ **do**
        **Calculate Sub-problem 2:**
        **for** $i \in N$ **do**
            **update** $UE_i = Rand(UE)$
        **end for**
    **end for**
**end for**

the dynamic algorithm (Algorithm 2). Pseudo-code shown in Algorithm 3 describes Sub-problems 1 and 2 of the heuristic algorithm in detail. In line 2, the essential parameters required in the algorithm are defined. The output of Sub-problem 1 is the heuristic cell-to-TAL/MME assignment. In lines 4-14, the algorithm calculates the UE crossing rates based on their average speeds and then sets the value of $X_{ij}^{lst}$ that corresponds to the maximum crossing rates between the cells. Line 16 defines the input of Sub-problem 2. The output of Sub-problem 2 is the usage ratio of the lists in each cell. In lines 18-30, the usage ratio of each list to which a cell belongs is determined by performing a search of all the lists that serve the cells. The ratio is calculated by distributing the usage of the TALs/MMEs that serve the cells with equal percentages. Finally, in lines 31-37, the signaling overhead of both schemes is calculated using the appropriate equations.

**Algorithm 3** Heuristic Algorithm

**Sup-Problem 1**

**Input:** $lst = \{1, 2, ..., L\}$: All lists = MMEs
  $TA = \{1, 2, ..., N\}$: All tracking areas
  $UE$: Average number of UEs in cell $i$
  $v_{Range} = [v_0 : v_{max}]$
  $v = Random$ such that $v \in v_{Range}$
  $Ui$: Density of UEs in cell $i$
  $R$: Cell diameter
  $UE_i$: Initial number of UEs in cell $i \in TA$.
  $X_{ij}^{lst}$: Initial random cell–to-TAL/MME assignment

**Output:** $X_{ij}^{lst}$: Optimized cell-to-TAL/MME assignment

**define** $UE_r^{ij} = \frac{Ui.L.v}{\Pi}$: Average crossing rate from $i$ to $j$

**define** $Hij = \frac{UE_r^{ij}}{UEi}$

**for** $i \in N$ **do**
  **for** $lst \in L$ **do**
    **if** $UE_r^{ij}$ = MAX **then**
      **SET** $X_{ij}^{lst} = 1$
    **else**
      **continue**
    **end if**
  **end for**
**end for**

**Sup-Problem 2**

**Input:** $lst = \{1, 2, .., L\}$: All lists
  $TA = \{1, 2, ..., N\}$: All tracking areas
  $X_{ij}^{lst}$: Cell-to-list binary indicator

**Output:** $\sigma_i^{lst}$: Usage ratio of list $lst$ in cell $i$

**for** $i \in N$ **do**
  **define** $N_L = \phi$
  **for** $lst \in L$ **do**
    **if** $i \in lst$ **then**
      **update** $N_L = \{N_L \cup N\}$
    **else**
      **continue**
    **end if**
    **for** $N \in N_L$ **do**
      $\sigma_i^{lst} = \frac{1}{||N_L||}$
    **end for**
  **end for**
**end for**

**Calculate:**
$C_\rho(i)$ as per (3.6f)

**if** Centralized Scheme **then**
  **Calculate:**
  $C_{Tu}(i)$ as per (3.6d)
**else**
  **Calculate:**
  $C_{Tu}(i)$ as per (3.6e)
**end if**

**Calculate:**
$\sum_i^N Cu\rho(i) = \sum_i^N \big(C_\rho(i) + C_{Tu}(i)\big)$

### 3.4.2 Algorithm Complexity

To justify the need for the heuristic algorithm, we must first discuss the complexity of the decomposition algorithm. The original problem presented in (4.5) is intractable since the model is a mixed-integer nonlinear programming (MINLP) problem with quadratic equality constraints, which is a well-known NP-hard problem that is difficult to solve to optimality [42]. Therefore, the decomposition model has been adopted to solve it. Sub-problem 1 has $O(L^{\frac{N}{\kappa}})$ complexity because each list can have at most $\kappa$ cells among $N$ available cells with possible overlap. Sub-problem 2 has $O(LN^2)$ complexity because the usage ratio of each list in all the cells is calculated for each cell. Hence, the total complexity of the decomposition algorithm is $O(L^{\frac{N}{\kappa}}) + O(LN^2) = O(L^{\frac{N}{\kappa}})$. On the other hand, the heuristic algorithm has much lower complexity. Using the proposed heuristic algorithm, Sub-problem 1 has $O(LN)$ complexity because a decision as to whether each cell belongs to the list is taken in each iteration. Sub-problem 2 also has $O(LN)$ complexity because the usage ratio of each list within each cell is calculated. Therefore, the total complexity of the heuristic algorithm is $O(LN)$, which is much lower than that of the decomposition algorithm.

## 3.5 Performance Evaluation

To evaluate the performance of the proposed models, we performed **MATLAB** simulations at four different speeds, categorized as very slow (0–8 $m/s$), slow (8–16 $m/s$), normal (16–25 $m/s$), and fast (25–33 $m/s$). The decomposition and heuristic models were implemented and compared with each other. We note that each sub-problem has the same importance; thus, the weight factors $\alpha$ and $\beta$ were assigned values of 1. The simulation tested the models statically and dynamically in the case of the centralized and distributed schemes. The dynamic algorithm presented in Algorithm (2) was implemented in the decomposition and heuristic algorithms, whereas the same analogy provided in our previous chapter was used for the static algorithm. Moreover, a random algorithm was added in the evaluation for the purpose of providing a simpler approach to replace Sub-problem 1, as it generates random cell-to-TAL/MME allocation dynamically. Moreover, the random technique enables us to evaluate the effectiveness of frequent random allocation without prior knowledge of the signaling overhead. This approach offers a less complex solution that does not affect the core network. The total signaling overhead of the TAU and paging combination was employed as a performance metric for the proposed algorithms. We used **MATLAB**'s built-in functions **intlinprog** and **linprog** to optimally solve Sub-problems 1 and 2, respectively. The simulation assumed an environment of 10 cells, each having a different number of UEs. The number of UEs was distributed uniformly among the cells with an average of 100 users.

### 3.5.1 Total Signaling Cost

Figure 3.2 shows the average total signaling overhead caused by paging and TAU at different speeds for the decomposition-based algorithms, namely the dynamic, random, and static algorithms, in the case of the centralized scheme. We note that the SON dynamic algorithm outperforms the static and random algorithms. On the other hand, the random optimal algorithm shows better performance in terms of the signaling overhead. This is because the random algorithm dynamically changes the TAL in a random manner, whereas the static algorithm maintains the same TAL at all times. We also note that the UE speed has a significant impact as it exceeds 25 $m/s$.

Figure 3.3 shows the average total signaling overhead at different speeds for the decomposition-based algorithms in the case of the distributed scheme. The results show the same trends as those of the centralized scheme. Furthermore, we can conclude that the centralized scheme outperforms the distributed scheme because of the frequent MME relocation in the latter case.

A comparison between the decomposition and heuristic algorithms is shown in Figures 3.4 and 3.5. Figure 3.4 shows the comparison in the case of centralized scheme, where it is observed that the heuristic approach shows acceptable behavior at lower speeds compared to the decomposition model. This is because the decision variable sigma diversifies the TAL/MME pool, thereby minimizing the TAU signaling and the overall signaling cost. Figure 3.5 shows a comparison between the decomposition and heuristic methods in the case of distributed scheme. The heuristic algorithm offers a near-optimal solution compared to the optimal decomposition model. This is a result of the dominant value of the MME relocation weight, which has a major effect on the signaling in the case of the distribution scheme.

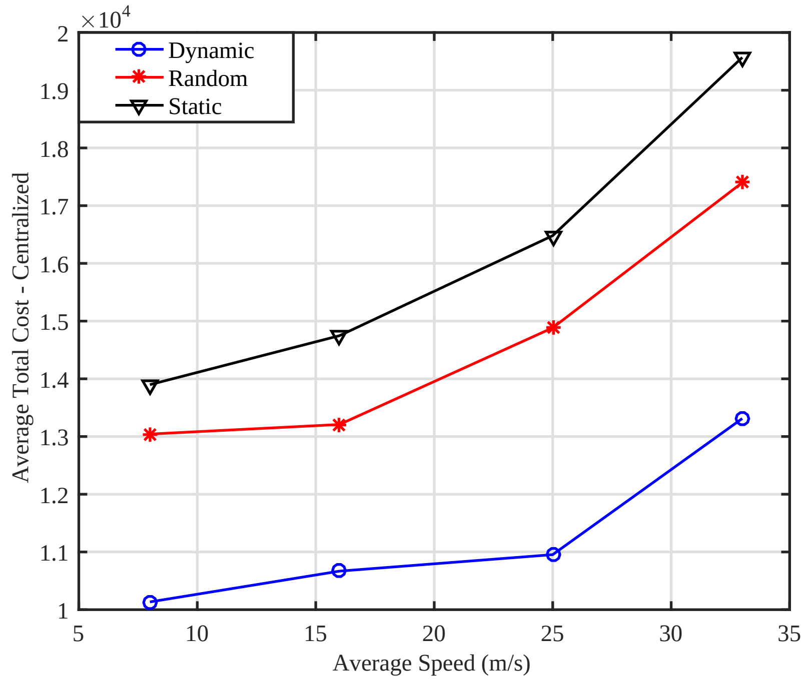


Figure 3.2: Total signaling overhead cost for the centralized scheme

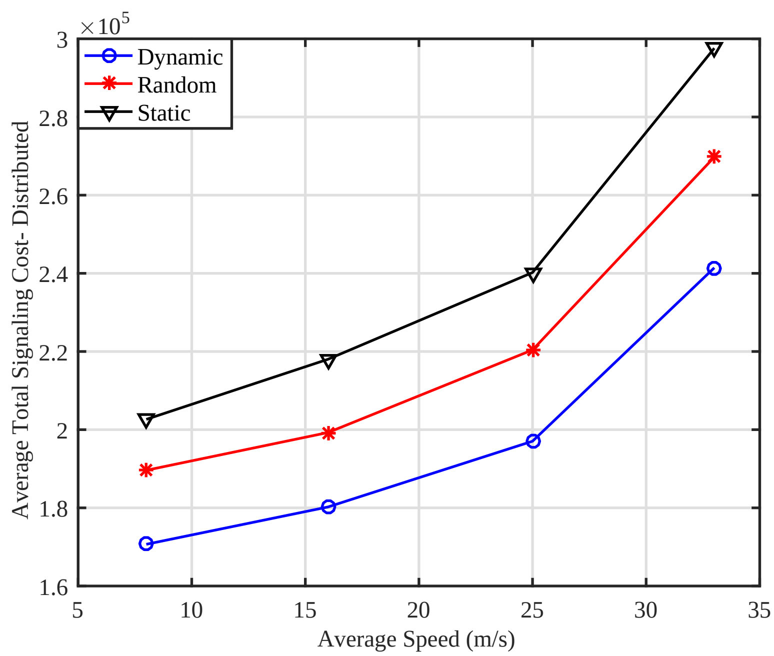


Figure 3.3: Total signaling overhead cost for the distributed scheme

Table 3.2 summarizes the average percentage improvement in the random and dynamic assignment algorithms compared to the static algorithm. In the centralized case, the random algorithm shows an improvement of 8%–16.45%, while the dynamic algorithm shows an improvement of 25.23%–34.98%. Similarly, the random algorithm results in an improvement of 6.53%–9.37% and the dynamic algorithm results in an improvement of 15.74%–18.88% in the distributed case.

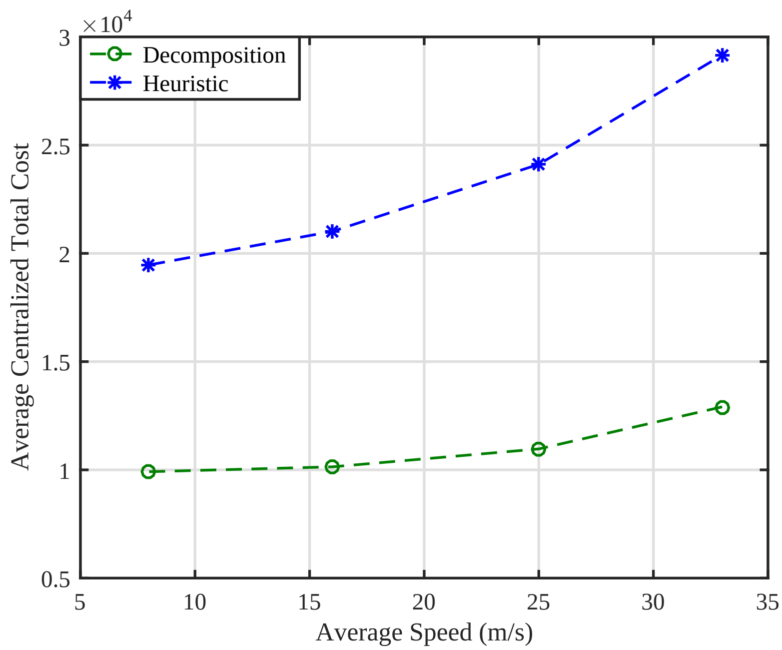


Figure 3.4: Total signaling overhead cost in the centralized scheme for decomposition algorithm vs. heuristic algorithm

Table 3.2: Average Improvement Percentage

| **Parameter** | **Random** | **Dynamic** |
|---|---|---|
| **Total Signaling Cost-Centralized** | [+8%, +16.45%] | [+25.23%, +34.98%] |
| **Total Signaling Cost-Distributed** | [+6.53%, +9.37%] | [+15.74%, +18.88%] |
| **Power Consumption** | [+16.6%, +17.7%] | [+28.83%, +39.32%] |

### 3.5.2 Power Efficiency

In this subsection, we discuss the efficiency of the proposed dynamic approach in terms of power consumption. One way to measure the efficiency of the SON dynamic

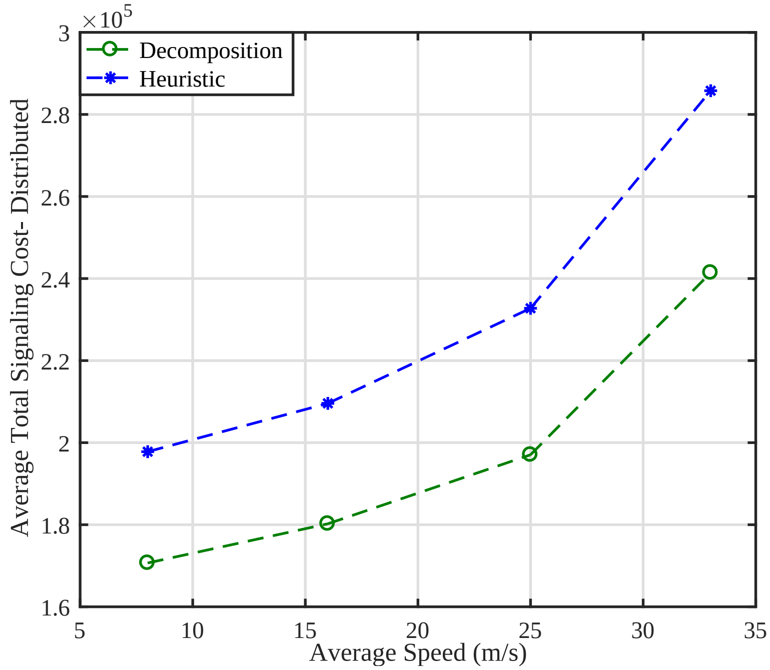


Figure 3.5: Total signaling overhead cost in the distributed scheme for decomposition algorithm vs. heuristic algorithm

model is to evaluate the UE battery life. It is estimated that each TAU procedure consumes around 10 mW of a regular smart-phone battery. The results in Figures 3.6 and 3.7 depict the average total battery consumption for one UE in an hour. Figure 3.6 compares the dynamic, random, and static approaches. The figure indicates significant power savings in the dynamic approach compared to the static approach. In addition, the random dynamic approach shows slightly lower power consumption than the static approach. Figure 3.7 compares the dynamic decomposition and heuristic solutions. It is clear that the decomposition solution performs better. However, the heuristic solution offers near-optimal results at slow speeds. Table 3.2 summarizes the power savings in the random and dynamic assignment algorithms compared to that in the static algorithm. The power savings range from 16.6% to 17.7% for the random algorithm and from 28.83% to 39.32% for the dynamic algorithm.

### 3.5.3 Related Work vs. Our Approach

This subsection compares the latest related methods in the literature with our approach. Three studies have been considered in the comparison: [9],[7], and [37]. The

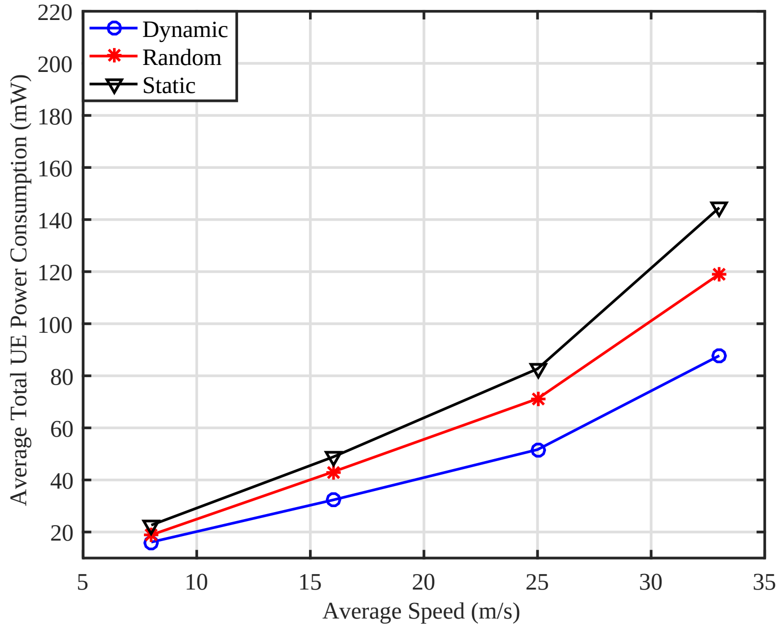


Figure 3.6: Average total UE power consumption (mW)

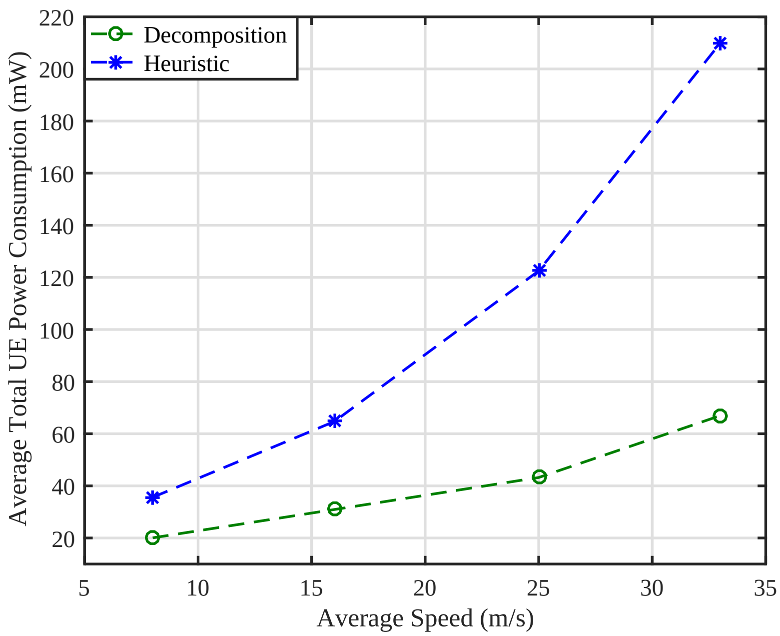


Figure 3.7: Average total UE power consumption (mW) for decomposition algorithm vs. heuristic algorithm

authors proposed three different approaches that dynamically solve the problem of cell-to-TA/TAL assignment for the purpose of minimizing the total signaling overhead due to TAU and paging signals. Specifically, two of the studies, i.e., [9] and [37], have proposed a SON mechanism to enhance the intelligence of their algorithms. However, we note that none of the aforementioned studies have included MME realization in their solutions, which is an important factor in the signaling overhead. Consequently, both MME realizations are adapted in the compared studies in order to ensure accurate and fair comparison. All the compared studies use the ring-based approach for cell-to-TAL assignment.

Figures 3.8 and 3.9 show the total signaling overhead in the centralized and distributed schemes. The key difference between our technique and other related methods is that our algorithm uses a smart technique for choosing the candidate cells to be assigned in TAL, whereas the other methods use the conventional cell-to-TAL assignment [7],[9], [37]. The smart selection technique outperforms the ring-based techniques because the smart selection assignment of cells alleviates the frequent TAU signaling by decreasing the probability of the UE moving from one cell to another cell that is not in the same tracking area list. Furthermore, the smart selection technique includes a greater number of cells that are more likely to be visited by the UE; thus, it is very efficient, especially for UEs moving at high speeds. Another key metric for the variation in the results is the overlapping TAL, which increases the probability of fewer TAU updates caused by the UE when it travels from one cell to another. Finally, the method for triggering the dynamic configuration or SON technique is also an important factor that affects minimization of the signaling overhead. The SON technique is used in [9], and it facilitates the transition between stages through timers and an activation threshold for triggering the dynamic configuration. The authors statically solved the problem of cell-to-TAL assignment until the activation threshold was reached. Although the activation threshold is not defined clearly, the technique used is too complicated to be implemented in a real-world scenario, especially with the rapid variation in UE velocity. Another SON technique was proposed in [37], where the authors introduced an overlapping TAL scheme solved statically in a manner similar to [9] until a threshold was reached. The main difference in this study is the introduction of the overlapping TAL technique, which offers the advantage of lower TAU signaling overhead. However, no available SON technique for solving TAL assignment has adapted the triggering scheme based on the velocity variation related to the UEs. Our scheme continuously adapts the TAL assignment based on the average UE velocity.

Figure 3.10 shows the power consumption of the related methods and our method. The comparison has been made at different speeds. Our method achieves greater power savings, which vary from +32.7% to +46.9%.

Table 3.3 summarizes the given comparison in terms of the average improvement

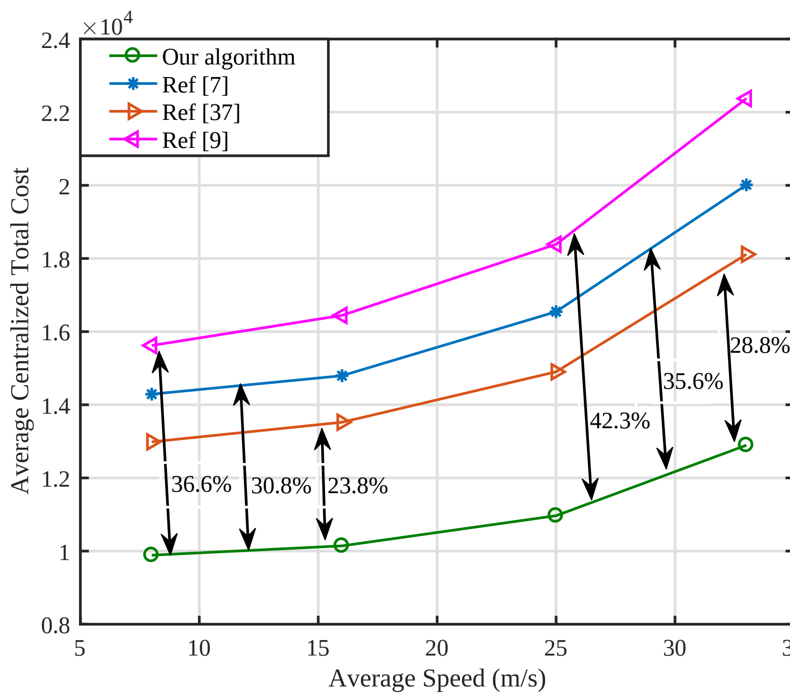


Figure 3.8: Total signaling overhead cost comparison between our algorithm (centralized) and related methods

percentage between our method (SON Dynamic/Smart) and the related methods. Our method outperforms the related methods significantly in terms of the centralized and distributed schemes as well as in terms of power consumption. For instance, the SON Dynamic/Smart algorithm in the case of the centralized scheme outperforms the method proposed in [37] by +23.87% to +28.82% which uses overlapping and SON techniques. The lower percentage is taken in the case of the lowest speed while the higher percentage is taken in the case of the highest speed, as shown in Figure 3.8.

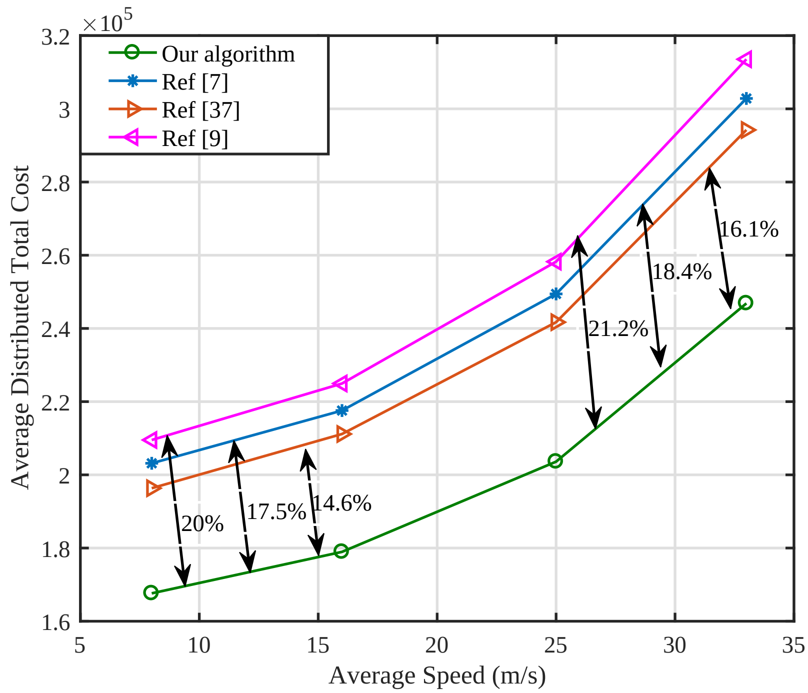


Figure 3.9: Total signaling overhead cost comparison between our algorithm (distributed) and related methods

Table 3.3: Average Improvement Percentage Compared With Related Methods

| **Algorithm Used** | **Overlapping** | **SON-Enabled** | **SON Dynamic/Smart Algorithm Centralized** | **SON Dynamic/Smart Algorithm Distributed** | **Power Consumption** |
|---|---|---|---|---|---|
| **Ref [7]** | $\surd$ | $\times$ | [+30.8%, +35.69%] | [+17.51%, +18.49%] | [+36.48%, +39.85%] |
| **Ref [37]** | $\surd$ | $\surd$ | [+23.87%, +28.82%] | [+14.62%, +16.11%] | [+28.73%, +32.66%] |
| **Ref [9]** | $\times$ | $\surd$ | [+36.69%, +42.37%] | [+20%, +21.27%] | [+42.84%, +46.91%] |

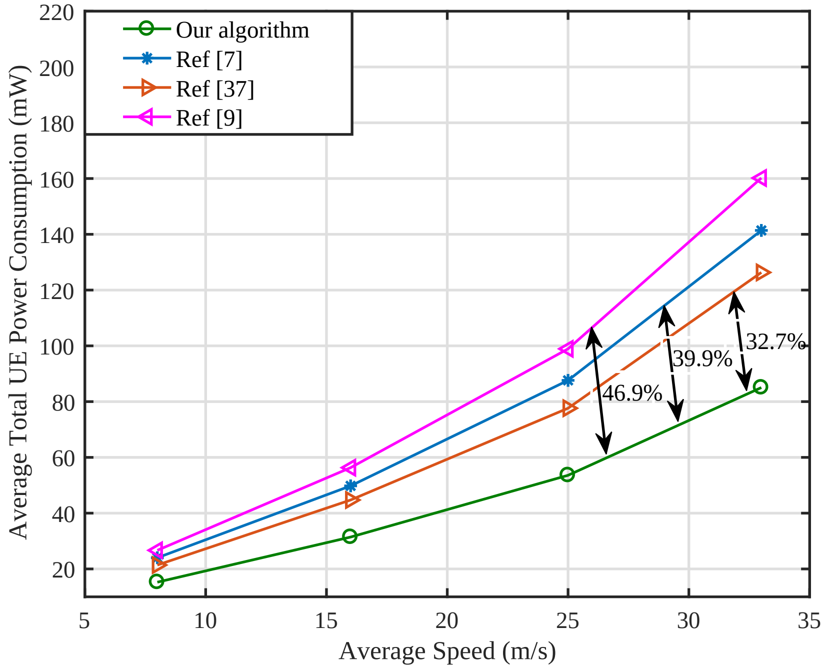


Figure 3.10: Average total UE power consumption (mW) comparison between our algorithm and related methods

## 3.6 Conclusion

SON, a new concept introduced in release 8 of 3GPP for LTE networks, is a promising paradigm that enables self-planning, self-managing, and self-optimizing of networks. Therefore, SON has been widely accepted across different applications. A number of SON use cases have been proposed to overcome increasing operational expenses. In this context, we proposed a SON approach to alleviate the signaling overhead caused by TAU and paging. We used SON as an enabler to perform dynamic cell-to-TAL/MME reconfiguration. Our approach can be considered as a SON use case that significantly minimizes the signaling overhead. Furthermore, two schemes used in the previous work, namely the centralized and distributed MME pooling schemes, were implemented and investigated dynamically. We used the well-known fluid flow model to simulate the movement of UEs within the system. The model consists of two sub-problems derived from our original NP-hard problem formulation. The first sub-problem is a binary integer programming problem, whereas the second is a linear programming problem. In addition, a new smart selection method was proposed to intelligently select the potential cells in the TAL/MME. Our method was shown to outperform conventional ring selection, which is commonly used in the literature. Finally, a less complex heuristic solution was proposed, which is easy to implement and gives a sub-optimal result. The results showed that the dynamic decomposition solution also achieves greater power saving than previous methods.

# Chapter 4
# Optimal Location Management in LTE Networks using Evolutionary Techniques

## 4.1 Introduction

In this chapter, the signaling overhead caused by the location techniques is studied. The concept of TAL is implemented and optimized using the evolutionary artificial optimization techniques. The goal is to minimize the total signaling overhead caused by TAU and paging. To the best of our knowledge, the proposed study is the first to include evolutionary artificial techniques in optimizing the location management techniques. Three techniques are implemented namely; Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), and Gravitational Search algorithm (GSA).

In recent years, PSO, ABC, and GSA have been considered to be reliable and promising evolutionary techniques for locating the global optima for a variety of optimization problems in different contexts with rapid rate of convergence as seen in different applications [2, 4, 10, 11]. In general, the optimization technique converges to the optimal solution in the search space with a predefined size where the size is defined through a number of optimized parameters satisfying some constraints. Evolutionary algorithms differ from one to another such that no specific technique could achieve best results for all optimization problems. Furthermore, some heuristic techniques may converge to the optimal solution faster than the others for specific problems and slower for other problems. Therefore, the three selected algorithms belong to different heuristic search families. For example, PSO relies on swarming and nature behavior of birds [1], ABC stands on colony structure and the main role of three main groups within the colony [3] and GSA is based on gravity laws and mass interactions between particles [5]. The obtained simulation results illustrate the effectiveness of the proposed approach in finding minimum total signaling overhead resulted in paging and tracking area update. Moreover, the applied optimization techniques offer an efficient solution for large scale problems which need fast convergence. Lastly, the results show that the power consumption of the user equipment hand-held devices is minimized as well.

The remainder of this chapter is organized as follows. Section 4.2 reviews studies on mobility management techniques. Section 4.3 presents a detailed description of the system model and discusses the problem definition, design hypothesis and problem formulation. Section 4.4 describes the used heuristic evolutionary algorithms to solve the optimization problem. Section 4.5 discusses the performance evaluation and results. Finally, Section 4.6 concludes the chapter.

## 4.2 Related Work

The previous two chapters discussed the TAU and paging signaling overhead with regards to TAL construction with the perspective of relating the management mobility entity with TAL. Both the studies proposed a centralized and a distributed MME scheme. The centralized scheme relates each TAL construction to one MME, whereas the distributed scheme relates each TA which represents each cell to one MME. Both the studies explore the significance of having different distribution of MME to the whole system and the total signaling overhead effect. However, the second chapter proposed a static technique that distributes the cells to tracking are lists in fixed time. It is shown that the centralized technique offers a more optimized signaling overhead when compared to the distributed one. Whereas the third chapter proposed an intelligent dynamic technique that redistributes the cells among the tracking area lists within a defined time calculated by a proposed smart selection scheme. The smart selection scheme helps including the cells that are more probable to have large handover between each other in the same list. The study showed a significant improvement in optimizing the signaling overhead.

Various heuristic techniques have been proposed to solve the problem of location management in different technologies. In [44] the author has proposed a solution for partitioning the location area in GSM based on the Genetic Algorithm (GA). The proposed solution seeks to avoid the bad resource utilization in terms of bandwidth along with increasing the efficiently of cells partitioning. However, the proposed solution was initial and did not cover most of the problem constraints such as the number of the cells in the location area, the dimension of the cells, and the mobility behavior of the users. Similar study that uses a modified GA to solve the problem of finding an optimal location areas configuration of wireless network has been proposed in [45]. The study introduced several modifications in the Genetic Algorithm properties specifically in the mutation operation. The authors argued that the design of an optimal location area configuration relies on different parameters related to the properties of the location area. In [46] the same authors proposed another solution that jointly uses the GA with the Hopfield neural network for finding the optimal configuration of location areas. The hopfield neural network is used to expedite the convergence of the solution.

A hybrid GA and PSO solution has been proposed in [47] for solving the minimization of local management overhead cost. Both algorithms have combined to improve the quality of the solution and increase the speed of convergence. The proposed solution allows crossover operation that can improve the diversity of fitness values. The authors have shown that the proposed solution has much better successful convergence rate than other solutions proposed in the literature.

An integer programming model has been presented in [48] to solve the TA reconfiguration problem. The study offered a dynamic reconfiguration procedure which has been compared to the performance of Tabu search and genetic algorithms. The proposed algorithm, however, has high complexity in spite its ability to outperform the tabu search and genetic algorithms and cannot be applied in the large-scale network.

Few studies have used the evolutionary optimization techniques in some LTE network applications. Authors in [49] proposed a joint particle swarm optimization and genetic algorithm to solve a resource allocation problem for the sake of maximizing the throughput for the user under the minimum rate requirements and the maximum transmission power constraints. The two techniques have been jointly used to enhance the speed of the convergence while finding the optimal solution.

A more detailed study by the same authors which follows the same approach was addressed in [50]. The proposed study uses the same techniques to solve the radio resource allocation problem in order to maximize the energy efficiency in the system under the quality of service constraints. The problem consisted of three sub-problems namely: resource blocks allocation, power distribution, and modulation schemes assignment problem.

A different problem related to the propagation model was tackled using the particle swarm optimization algorithm in [51]. The authors developed an online propagation correction model based on users' measurement information. The developed model helps improving the planning of the LTE network under the realistic conditions by analyzing the diverse behavior of the mobile user equipment.

## 4.3 System Model

### 4.3.1 Problem Definition

The objective of this study is to minimize the total signaling overhead caused by paging and TAU. Two studies have been proposed earlier in chapter two and three, which implement various methods to minimize the total overhead. Chapter two proposed a study that minimizes the overhead gain by using two pooling schemes namely; centralized and distributed. Both approaches relate the TAL distribution to MME in the system. It is found that the centralized scheme outperforms the distributed one. Chapter three presents a Self Organizing dynamic technique that can minimize the total signaling overhead instantaneously. This chapter differentiates from the previous ones by implementing evolutionary artificial optimization techniques. Three techniques are used, each of which is implemented to minimize the total overhead caused by paging and TAU. The optimization techniques used allow implementing the optimization in a large scale network. It is worth mentioning that only the centralized scheme is used since it has better performance than the distributed algorithm in terms of minimizing the total signaling overhead.

### 4.3.2 Paging and TAU Trade-off

Paging and TAU are the essential methods for locating idle users within the cellular network. The signaling resulted by TAU takes place whenever a user travels from cell to another that is not within the same TAL. On the other hand, paging signaling takes place whenever a user need to be located by the core network. The signaling cost of both paging and TAU is directly dependent on the size of TAL. In other words, the relation between TAU and paging is inversely proportional and dependent on the size of TAL. Therefore, the higher the number of cells that are assigned to TAL, the lesser the signaling overhead caused by TAU and the more signaling overhead caused by paging and vice versa. This is due to the large number of cells that are accommodated within the lists in which yields to less probable users travel from one cell to another that is not within the same list. On the contrary, paging signaling overhead would increase because of the large number of cells that would be paged to locate a certain user.

### 4.3.3 Design Hypothesis

As stated previously, the centralized MME pooling scheme is used in this context which allocates each TAL to an individual MME element. In this model, a TA would represent one cell as this simplifies the problem and avoids the cell-to-TA assignment to impose additional complexity. This is because of the fact that Cell-to-TA reallocation yields to service interruption. Moreover, it must be mentioned that in this context, cell and TA can be used interchangeably and will have the same constraints. Figure 4.1 shows the centralized MME pooling schemes.

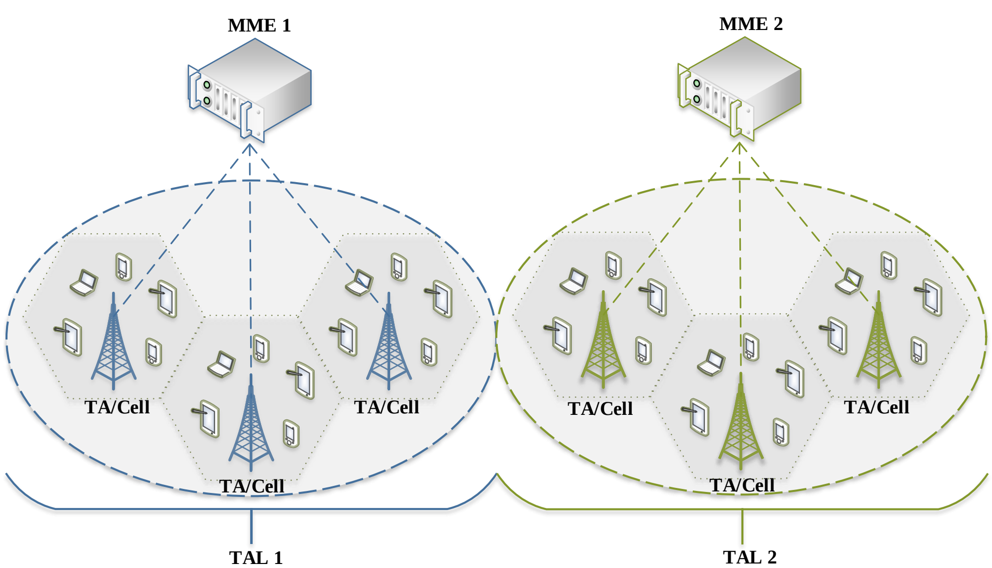


Figure 4.1: Centralized MME pooling scheme

The model presents two tiers of assignment: The first handles the Cells-to-TALs/MMEs assignment that takes place in the core network, whereas the second tier handles the TALs/MMEs to UEs assignment for each cell in the system. In terms of Cells-to-TALs/MMEs assignment, the system calculates the mobility pattern of the UEs within the cells and assigns or reassigns the cells inside the TALs/MMEs with the purpose of minimizing the TAU and paging signaling overhead. On the other hand, TALs/MMEs-to-UEs assignment, each cell distributes a number of TALs to a portion of UEs by an estimated usage ratio known as $\sigma_i^{lst}$. Variable $\sigma_i^{lst}$ is a decision variable that determines the usage ratio of each TAL/MME at each cell. This enables an overlapping of TALs among the cells thus facilitating more variety of cells to be accommodated within the lists. Table 4.1 summarizes the important notations used in the system model.

Table 4.1: Table of Notations-Part I

| | | |
|---|---|---|
| $C_{Tu}(i)$ | : | Total signaling cost of TAU in cell $i$. |
| $C_u$ | : | TAU cost of UE moving from one list to another. |
| $UE_i$ | : | Total number of UEs served by cell $i$. |
| $\rho$ | : | Paging arrival rate. |
| $H_{ij}$ | : | Probability that a user moves from cell $i$ to cell $j$. |
| $C_\rho$ | : | Paging cost of particular user equipment. |
| $C_{T\rho}(i)$ | : | Total paging cost in cell $i$. |
| $Cu\rho(i)$ | : | Total paging and TAU overhead in cell $i$. |
| $L$ | : | Total number of lists. |
| $N$ | : | Total number of cells. |
| $lst$ | : | Individual list where $lst$ is a subset of $L$. |
| $\kappa$ | : | Maximum number of TAs that are assigned to list $lst$. |
| $\omega$ | : | Cost of MME relocation during handover. |
| $HX^{lst}(i)$ | : | Inter-list handover rate of users in cell $i$. |

Table of Notations-Part II

<u>Decision Variables</u>

| | | |
|---|---|---|
| $\sigma_i^{lst}$ | : | Usage ratio of each list $lst$ in cell $i$. |
| $X_{ij}^{lst}$ | = | $\begin{cases} 1, & \text{if cells } i \text{ and } j \text{ belong to } lst, \\ 0, & \text{otherwise} \end{cases}$ |

<u>Optimization Variables</u>

| | | |
|---|---|---|
| $x$ | : | position of particle, food source or agent. |
| $v$ | : | velocity of particle or agent. |
| $a$ | : | agent acceleration. |
| $J_1$ | : | Cost function 1 to be minimized. |
| $J_2$ | : | Cost function 2 to be minimized. |
| $\mathcal{J}$ | : | Total cost function. |
| $\mathcal{N}$ | : | Population size. |
| $\mathcal{P}$ | : | Number of parameters to be optimized. |
| $P$ | : | Probability of food source. |
| $G$ | : | Gravitational constant. |
| $\mathcal{M}$ | : | Mass of agent. |
| $\mathcal{F}$ | : | Gravitational force. |
| $\mathcal{R}$ | : | Euclidean distance. |

#### 4.3.3.1 Cell/TA-to-TAL/MME Assignment

The model seeks to determine the optimal assignment of cell/TA-to-TAL/MME in centralized pooling scheme. Each TAL represents a particular MME as a one-to-one basis. In this study, the assignment of Cells-to-TAL/MME is defined by a binary decision variable $X_{ij}^{lst}$ which indicates whether cells $i$ and $j$ reside in the same tracking area list.

As an example of the search space the decision variable $X_{ij}^{lst}$ can have, let us assume, the maximum number of cells/TAs in a list is 2 and that there are 3 lists or MMEs. Therefore, there will be 2 x 3 possible combinations of the binary variable $X_{ij}^{lst}$ generated in the system. The following matrix denoted as $X^1$ shows an example of the choices that list 1 can have:

$$X^1 = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array} \tag{4.1}$$

The example shows that cells 2 and 3 are in list one so the values of $X_{22}^1$, $X_{33}^1$, $X_{23}^1$ and $X_{32}^1$ are set to 1 whereas the rest is set to zero. From the aforementioned information, the number of cell-to-TAL combinations can be generalized as $L \times N$ and can be represented in the following matrix:

$$\begin{bmatrix} X_{11}^1 & X_{12}^1 & \dots & X_{1N}^1 \\ X_{21}^1 & X_{22}^1 & \dots & X_{2N}^1 \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1}^1 & X_{N2}^1 & \dots & X_{NN}^1 \end{bmatrix} \dots \begin{bmatrix} X_{11}^L & X_{12}^L & \dots & X_{1N}^L \\ X_{21}^L & X_{22}^L & \dots & X_{2N}^L \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1}^L & X_{N2}^L & \dots & X_{NN}^L \end{bmatrix}$$

Another vital aspect related to the assignment of cells to TALs is the maximum number of cells within the list. The following Equation 4.2 is to limit the number of cells within the list where the constant $\kappa$ is the maximum number of TAs in the list:

$$\sum_{i}^{N} \sum_{j, j \neq i}^{N} X_{ij}^{lst} \leqslant \kappa, \quad \forall lst \in L \tag{4.2}$$

In addition to limiting the number of cells, a different constraint is needed in order to eliminate the redundancy of cell/TA assignment in each TAL as shown below:

$$X_{ij}^{lst} = X_{ji}^{lst}, \quad \forall lst \in L, \ \forall i, j, i \neq j \in N \tag{4.3}$$

#### 4.3.3.2 TAL-to-UE

The assignment of the lists to the users takes place through each cell. As mentioned earlier, each TAL is associated to each MME than can be assigned to more than one cell. This can be modeled by the following constraint:

$$\sum_{ij,i\neq j}^{N} X_{ij}^{l} \leqslant \kappa, \quad \forall\, l \in L \tag{4.4}$$

### 4.3.4 Problem Formulation

The original model that was proposed in chapter two is used as a base model for both the centralized and the distributed pooling schemes. In this context, the formulation of the problem considers the centralized scheme owing to its significant performance advantage over the distributed scheme. The objective of the model is to minimize the signaling overhead resulting from both TAU and paging. The model is modified and divided into a bi-objective minimization problem as follows:

- **Objective function**

$$\min \mathcal{J} = \min \ (\alpha J_1 + \beta J_2) \tag{4.5a}$$

- **Cost functions**

$$Cu\rho(i) = C_{T\rho}(i) + C_{Tu}(i), \ \ \forall\, i \in N \tag{4.5b}$$

$$J_1 = \sum_{i}^{N} Cu\rho(i) \tag{4.5c}$$

$$HX^{lst}(i) = \sum_{lst=1}^{L} UE_i.H_{i,j}.(1 - X_{ij}^{lst}) \tag{4.5d}$$

$$J_2 = \sum_{i}^{N} HX^{lst}(i) \tag{4.5e}$$

- **Total TAU cost**

$$C_{Tu}(i) = UE_i.H_{ij}.C_u\Big[\sum_{lst=1}^{L} \omega.\ O^{M}_{lst}.\ \sigma_i^{lst}(1 - X_{ij}^{lst})\Big], \quad \forall\, i, j, i \neq j \in N \tag{4.5f}$$

- **Total Paging cost**

$$C_{T\rho}(i) = \rho\ .C_\rho\Big[\sum_{lst=1}^{L} UE_i.\sigma_i^{lst} + \sum_{lst=1}^{L}\sum_{j,i\neq j}^{N} UE_j.\ X_{ij}^{lst}\ .\sigma_j^{lst}\Big], \quad \forall\, i \in N \tag{4.5g}$$

- **Constraints**

$$\sum_{l=1}^{L} X_{ij}^{lst}.\sigma_i^{l} = 1, \quad \forall\, i \in N \tag{4.5h}$$

$$\sum_{ij,i\neq j}^{N} X_{ij}^{lst} \leqslant \kappa, \quad \forall\, lst \in L \tag{4.5i}$$

$$X_{ij}^{lst} = X_{ji}^{lst}, \quad \forall lst \in L,\ \forall\, i, j, i \neq j \in N \tag{4.5j}$$

$$0 \leq \sigma_i^{lst} \leq 1 \tag{4.5k}$$

$$O^{M}_{lst} \in 0, 1 \tag{4.5l}$$

$$X_{ij}^{lst} \in 0, 1 \tag{4.5m}$$

The objective function (4.5a) denoted as $\mathcal{J}$ is divided into two parts. The first part denoted as $J_1$ minimizes the signaling overhead caused by TAU and paging; the second part denoted as $J_2$ minimizes the inter-list handover. Both parts are weighted with factors represented as $\alpha$ and $\beta$. These weights can be controlled by the service provider as parameters to prioritize an objective value over the other. The first part of (4.5a) contains two cost functions, as shown in (4.5b), which combines the cost function of TAU as in (4.5f), along with the second term, which has the paging cost function, as in (4.5g). The tracking area update signaling overhead of each cell is determined by first calculating the number of UEs that reside in a cell $i \in N$ which have the probability to move to another cell that is not within the same list, and then multiplying that number by the cost of inter-MME reallocation ($\omega$) and the usage ratio of each list ($\sigma_i^{lst}$) in cell $i$. The paging cost function (4.5g), on the other hand calculates the signaling load of paging messages that are triggered by the UEs inside each cell $i$ served by each list. The cost considers the percentage of the overlapping lists used in cell $i$ multiplied by the decision variable $X_{ij}^{lst}$ which determines whether a cell $i$ and the neighboring cell $j$ belong to the same list. The second part of (4.5a) considers estimating the inter-list handover rate of the average number of UEs in a cell $i$, as given in (4.5d). Constraint (4.5h) ensures fair usage by a set of cells for every list/MME. Moreover, the load of MMEs is balanced throughout the cells under each assigned MME. Constraints (4.5k), (4.5l), and (4.5m) are the boundary constraints.

### 4.3.5 Mobility Pattern Model

We use the fluid flow model to simulate the mobility behavior of the users in the system. The fluid flow model is a well-known model that is commonly used in order to simulate the mobility pattern of the users. The model depicts the traffic flow rates of UEs moving out of a closed region represented as a cell or base station. For a certain cell $i$ with perimeter $L$, UE density $Ui$, and average UE velocity $v$, the average number of cell crossings per unit time is calculated as follows:

$$\frac{Ui.L.v}{\pi} \tag{4.6}$$

In this context, cells are hexagonal in shape with side length $R$; hence, $L$ is replaced with $6R$.

## 4.4 Evolutionary Optimization Techniques and Implementation-4

The following three subsections give a brief description of the three optimization techniques involved in objective function minimization of the proposed algorithm. The objective function is defined as a minimization problem to minimize the total signaling overhead caused by TAU and paging. Table 4.2 shows a brief summery of the techniques that are used in this study,

Table 4.2: A brief summery of ABC, GSA, and PSO techniques

| Metric | Technique Name | | |
|---|---|---|---|
| | ABC | GSA | PSO |
| Based on | Nature life of bees in the colony | Newton's law of gravitations | Social behaviors of birds swarming and fish schooling |
| Complexity | High | Fair | Low |
| Advantages | High quality solution | Good convergence rate | Easy to handle the constraints |

### 4.4.1 Particle Swarm Optimization

Particle swarm optimization (PSO) is a stochastic optimization technique based on the social behaviors of birds swarming and fish schooling. It implements a set of particles in the space to find the optimal solution [1], [2]. Each particle in the space is a candidate solution with swarming velocity and position. The total number of particles in the space represents the whole population. The optimal solution is defined by its best position and the algorithm is employed to guarantee the swarming of particles in the proximity towards the optimal solution in order to increase the probability of fast convergence. Therefore, the updated velocity and position of each particle can be obtained from (4.7) and (4.8) respectively as follows:

$$\begin{aligned} v_{i,j}(t) = \alpha(t) v_{i,j}(t-1) + c_1 r_1 (x^{\star}_{i,j}(t-1) \\ - x_{i,j}(t-1)) + c_2 r_2 (x^{\star\star}_{i,j}(t-1) - x_{i,j}(t-1)) \end{aligned} \tag{4.7}$$

$$x_{i,j}(t) = v_{i,j}(t) + x_{i,j}(t-1) \tag{4.8}$$

with $j = 1, 2, \ldots, \mathcal{P}$ and $\mathcal{P}$ is the number of parameters to be optimized, $i = 1, 2, \ldots, \mathcal{N}$ and $\mathcal{N}$ is the population size. Each of the size and number of populations should be selected with respect that the best position is equal or very close to the optimal solution. $x^{\star}_{i,j}$ and $x^{\star\star}_{i,j}$ represent the local and global solutions respectively, $r_1$ and $r_2$ represent random numbers between 0 and 1, $\alpha(t)$ is decremental inertia factor and can be selected such that $\alpha(t) = 0.99\alpha(t-1)$. In general, the iterative solution starts by $t = 1$ and ends by $\mathcal{N}$ such that $t = 1, \ldots, \mathcal{N}$. The personal and social influence of each particle are defined by $c_1$ and $c_2$ respectively. Figure 4.2 depicts the graphical illustration of PSO algorithm.

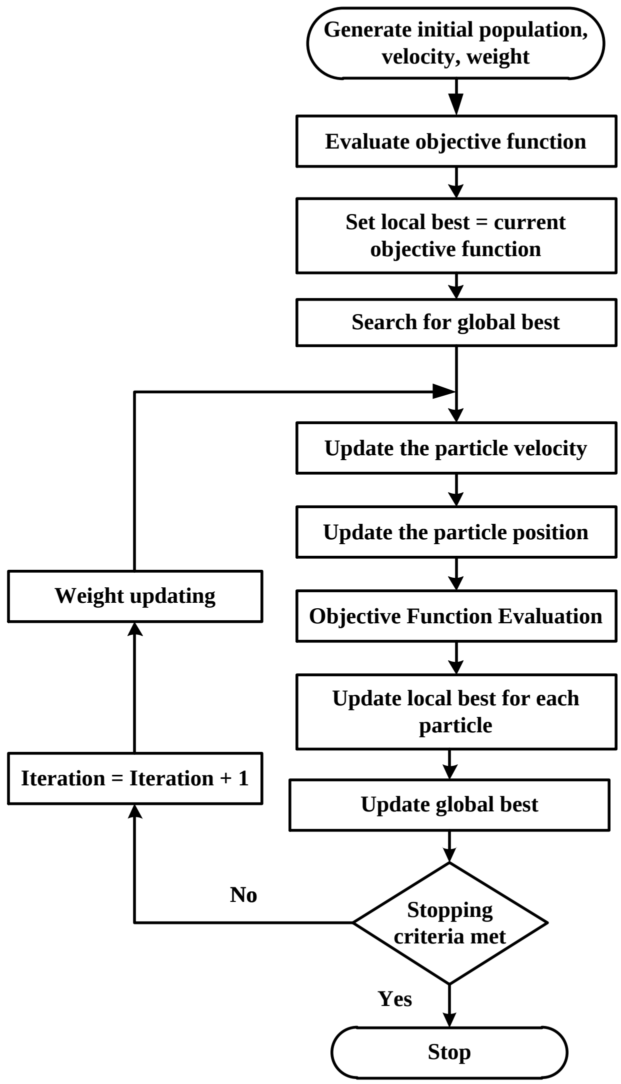


Figure 4.2: PSO computational flowchart [1, 2].

### 4.4.2 Artificial Bee Colony

Artificial Bee Colony (ABC) algorithm was first introduced in 2005 as a global optimization approach to mimic the natural life of bees in the colony [52]. The structure of colony is divided into groups namely; employed, onlooker and scout bees. The employed bees are responsible for searching randomly for the available sources (solutions) of food. This can be done by marking the position with best food as the best position. Once the position is marked, the employed bees dance to other bees to indicate the amount of nectar in the allocated source. Onlookers can differentiate between good and bad source of food by the duration of dance and the speed of shaking. The longer dance with higher speed of shaking refers to better food. It notes mentioning that scouts search for the source of food without referring to the food quality. Based on the nectar quality (solution's quality), scouts can be chosen to be employed bees and vice versa [52, 3]. Employed and onlooker bees are determined to find the optimal food source whereas scouts control the search process [3]. Food source position in the search space is modeled as follows:

$$x_{ij}^{new} = x_{ij}^{old} + \alpha(x_{ij}^{old} - x_{kj}) \tag{4.9}$$

The objective function is represented by the amount of nectar. The representation of the probability of onlooker bees for selecting a food source is depicted in the following:

$$P_i = \frac{\mathcal{J}_i}{\sum_{i=1}^{\mathcal{N}} \mathcal{J}_i} \tag{4.10}$$

where $i = 1, 2, \ldots, \mathcal{N}$ with colony size is defined by $2\mathcal{N}$, $j = 1, 2, \ldots, \mathcal{P}$ with $\mathcal{P}$ represents the number of parameters to be optimized, $j$ representisthe number of positions with $\mathcal{P}$ dimension, $\mathcal{J}_i$ is the associated objective function of $i$th, $k$ is a random number within the colony size where $k \in (1, 2, \ldots, \mathcal{N})$, $\alpha$ is a random number between 0 and 1. The ABC algorithm is illustrated as a flowchart Figure (4.3)

### 4.4.3 Gravitational Search Algorithm

Gravitational Search Algorithm (GSA) is a meta-heuristic technique based on Newton's laws of gravitations. It was first introduced in 2009 [5]. The algorithm follows the

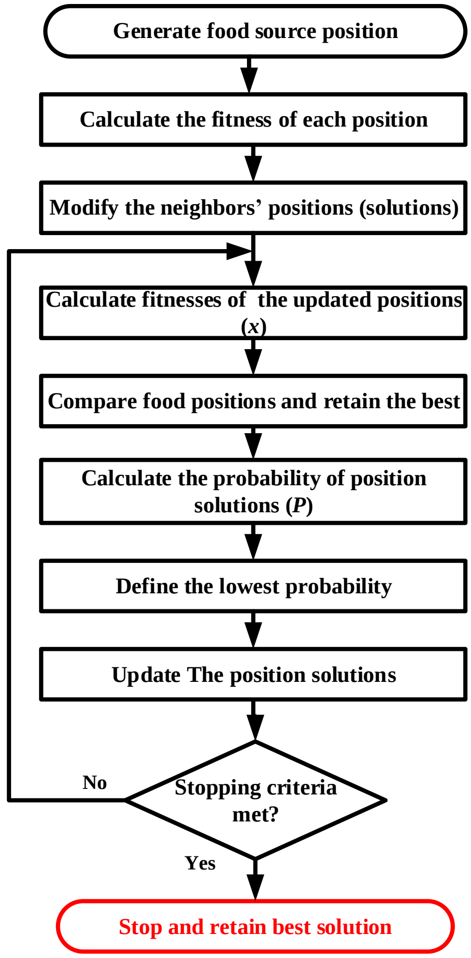


Figure 4.3: ABC computational flowchart [3, 4].

gravitational law: "for any two objects, every object is attracted to the other object by a force which is directly proportional to their mass and inversely proportional to their square distance". According to gravity principle, the gravitational force $\mathcal{F}$ can be defined between any two agents by:

$$\mathcal{F} = G\frac{\mathcal{M}_1\mathcal{M}_2}{\mathcal{R}^2} \tag{4.11}$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ represents masses of agent 1 and 2, $\mathcal{R} = ||X_i(t), X_k(t)||^2+\varepsilon$ represents the Euclidean distance between the two agents $i$ and $k$ with $X_i(t) = \left[x_{i,1}, \ldots, x_{i,\mathcal{P}}\right]$, $\varepsilon$ represents a small positive constant and $G$ represents the gravitational constant. $G$ is defined at any instant time $t$ or equivalently at the current iteration in the GSA algorithm by the following:

$$G(t) = G(t_0)\exp\left(-\frac{\beta t}{\mathcal{T}}\right) \tag{4.12}$$

with $G(t_0)$ is the initial value of gravitational constant at time $t_0 = 1$, $\beta$ is a positive constant and $\mathcal{T}$ is the final research time which can be assigned by number of iterations. The agent acceleration is given by:

$$a_{i,j}(t) = \frac{\mathcal{F}_{i,j}(t)}{\mathcal{M}_i(t)} \tag{4.13}$$

where $i = 1, \ldots, \mathcal{N}$ with $\mathcal{N}$ is number of agents and $j = 1, \ldots, \mathcal{P}$ with $\mathcal{P}$ is number of optimized agents within the agent. The force of particle $x_{i,j}$ is $\mathcal{F}_{i,j}$ and $\mathcal{M}_i$ is the mass of particle $i$. The velocity of each agent is:

$$v_{i,j}(t+1) = \alpha_i v_{i,j}(t) + a_{i,j}(t) \tag{4.14}$$

The range of velocity is correlated to the position boundaries by a factor $1/\lambda$ where $\lambda$ is called the speed division factor. Finally, the position is can be given as follows:

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \tag{4.15}$$

where $x_{i,j}(t+1)$ and $v_{i,j}(t+1)$ are the position and the velocity of parameter $j$ in agent $i$ at time $t+1$ or at the next iteration respectively. Also, $\alpha_i$ is a random number between 0 and 1. A small set $K_{best}$ is used to include the best solution through the search history with small size. After every iteration, the small set $K_{best}$ is updated if the algorithm find better results. A detailed explanation of the GSA is given in [5] and the flowchart in Fig (4.4) gives the synopsis of the algorithm [5, 53].

### 4.4.4 Implementation of the Optimization algorithms

For all the three algorithms, the assignment of Cells-to-TAL/MME $X^{lst}$ follows that $X^{lst} = \left[X^1, X^2, \ldots, X^L\right]$, $X^l$ is a square matrix with $n$dimension such that $[n, n] =$ size$\left(X^l\right)$ for all $l = 1, 2, \ldots, L$. According to (4.1), $X^l$ is symmetric and the non-zero

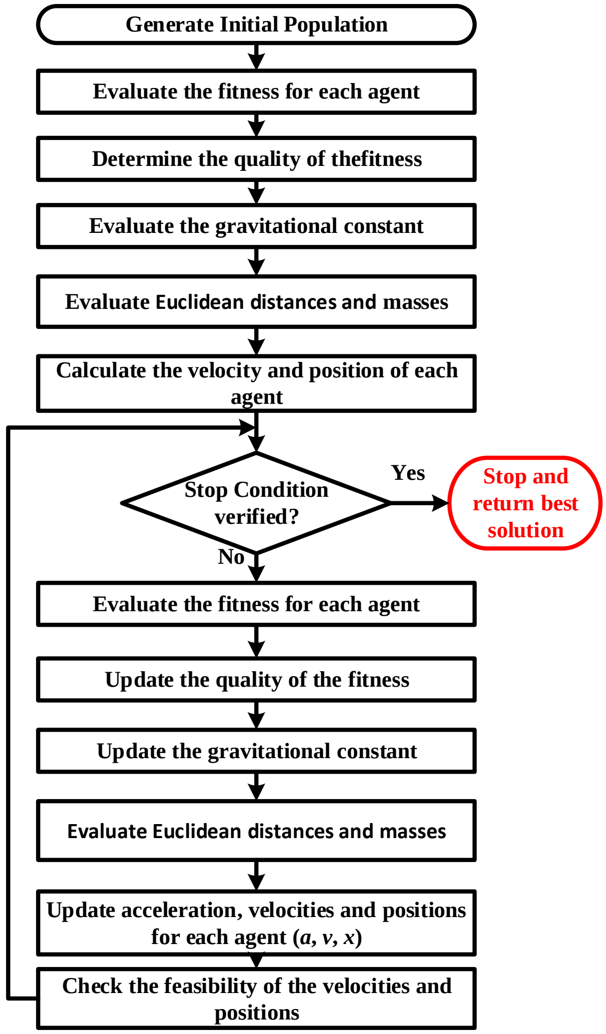


Figure 4.4: GSA computational flowchart [5].

row is repeated in a systematic order within list $l$. From the other side, the decision variable $\sigma^l$ is a row vector with $n$dimension. Also, the order of zero components in the non-zero row in $X^l$ matches the order of zero components in $\sigma^l$. Similarly, the order of one components in the non-zero row in $X^l$ matches the order of non-zero components in $\sigma^l$. Let's define a new variable $\bar{\sigma}^l$ with $n$dimension for all $l = 1, 2, \ldots, L$ such that $\bar{\sigma}^l$ is an auxiliary variable to be optimized. Hence, the total number of parameters to be optimized $\mathcal{P}$ is defined according to $\mathcal{P} = (n + n) \times L$ such that the first $n$ denotes the optimized parameters of non-zero rows in $X^l$, the second $n$ denotes the optimized parameters of the row vector $\bar{\sigma}^l$ and $L$ denotes number of lists. It should be noted that for each iteration, we have for any non-zero row $r$ in the list $l$

$$\begin{aligned} X_r^l &= x_{i,2(l-1)n+1:n(2l-1)} \\ \bar{\sigma}^l &= x_{i,n(2l-1)+1:2nl} \end{aligned} \tag{4.16}$$

where, $x_{i,j}$ be defined in each optimization algorithm. Next, we have

$$\sigma_j^l = \begin{cases} 0, & \text{if } X_{r,j}^l = 0 \\ \bar{\sigma}_j^l, & \text{if } X_{r,j}^l = 1 \end{cases} \tag{4.17}$$

$$\sigma^l = \frac{\sigma^l}{\sum\limits_j^N \sigma_j^l} \tag{4.18}$$

where $\sigma^l = \left[\sigma_1^l, \ldots, \sigma_N^l\right]$ for $j = 1, \ldots, N$. Optimization algorithms are employed to obtain the minimum objective function. Recall (4.5a)

$$\mathcal{J} = \alpha J_1 + \beta J_2 \tag{4.19}$$

where $J_1$ and $J_2$ be defined in (4.5) also $\alpha$ and $\beta$ are weighting factors which can be selected arbitrarily to give almost equal weights for both objectives. In fact, the proper selection of $\alpha$ and $\beta$ will enable the final solution to be very close to the best compromise solution between $J_1$ and $J_2$.

In the case of PSO algorithm, it can be easily deduced that (4.16) represents particles to be optimized $\mathcal{P}$ in each iteration $i$ for $i = 1, 2, \ldots, \mathcal{N}$ with $\mathcal{N}$ being the population size and the remaining steps can be easily followed from the flow chart in Figure 4.2, in addition to [1]. For the case of ABC algorithms, (4.16) denotes the position of the source food and (4.19) denotes the quality of the allocated food. In this case, $\mathcal{N}$ is the number of sources and it represents half the colony size. Finally, the current quality sources with size $\mathcal{N}$ are joined to the best sources that are obtained during previous iterations with size $\mathcal{N}$ to formulate the complete colony size $2\mathcal{N}$. Then, (4.10) is employed to provide ranking from best (minimum) to worst (maximum) sources (objectives). Now, we keep half of the colony size $\mathcal{N}$ which includes the best (minimum) sources (objectives) or more specifically, the colony with size $2\mathcal{N}$ is sorted in ascending order and we only keep the first half $\mathcal{N}$ and go for the next iteration process. For the implementation of GSA, (4.16) represents number positions and the rest of the steps are illustrated in the flow chart in Figure 4.4 [5].

## 4.5 Performance Evaluation and Simulations

In order to evaluate the performance of each techniques, a simulation is performed at different speeds categorized as very slow (0–8 m/s), slow (8–16 m/s), normal (16–25 m/s), and fast (25–33 m/s). The simulation is built with an environment that has a total number of 30 cells, each of which has an average of 100 users. The number of users is distributed in a uniform fashion among all the cells in the system. In addition the number of lists that are used in the system is 10. It is worth stating that the number of lists is also equivalent to the number of MME that are used in the system. Table 4.3 shows the configured simulation parameters values.

Table 4.3: Simulation Parameters & Values

| **Parameter** | **Value** |
|---|---|
| Number of cells $N$ | 10 |
| Average number of users | 100 per cell $i$ |
| Number of TAs | 10 |
| Paging rate $\rho$ | 0.05 |
| UE speeds | 0,8,16,25 and 33 m/s |
| Cell Radius | 500 m |
| TAU to Paging cost | 10:1 |

The simulation tests the convergence of each technique used in this study. There are 30 cells in the system where 16 cells are to be allocated within each list. The number of decision variables of the first stage is similar to the ones in the second stage. This is because the first stage is for optimizing the allocation of the candidate cells within each list whereas the second stage is for optimizing the procedure of overlapping the lists and MMEs within the cells themselves. This implies optimizing 600 decision variables resulting from the combination of 30 cells in each list for the 10 lists to be chosen from. The 600 decision variables comprise of 300 of $X_{ij}^{lst}$ and 300 of $\sigma_i^{lst}$. As such, the problem has been defined as an optimization problem and the minimum objective is attained through one of the well-known evolutionary techniques. In fact, each of PSO, ABC and GSA has been employed to obtain the minimum cost function and to compare the three techniques in terms of minimum cost and convergence rate. For fair comparison, we selected same values of each of $\mathcal{P}$ and $\mathcal{N}$ in the all three optimization algorithms. The setting parameters of PSO and GSA are presented in Table 4.4, whereas, there is no setting parameters for the ABC algorithm. The setting parameters in Table 4.4 were selected arbitrarily after set of trials.

Table 4.4: The Setting Parameters of optimization techniques

| | | | | | | |
|---|---|---|---|---|---|---|
| PSO | Parameter | $c_1$ | | $c_2$ | $\alpha\,(1)$ | |
| | Setting | 2 | | 2 | 0.99 | |
| GSA | Parameter | $\beta$ | $\lambda$ | $\varepsilon$ | $G(t_0)$ | $K_{best}$ |
| | Setting | 7 | 6 | 0.00001 | 1000 | 4 |

### 4.5.1 Total Signaling Cost

In this subsection, the total signaling of paging and tracking overhead is evaluated for each algorithm at each of the aforementioned speeds. The test is conducted to evaluate the convergence of the objective function $\mathcal{J}$ and its influence on the total signaling overhead. To guarantee the robustness of each optimization algorithm during the search process, each algorithm is employed to solve the problem five times considering random initialization.

The precision of convergence is measured through the standard deviation and means values. Furthermore, the uncertainty of different measurements is evaluated by the coefficient of variance or what is known as the Relative Standard Deviation (RSD).

Table 4.5 shows the minimum values of the objective function at different range of speed. The table depicts five experiments of the converged values along with the their mean and standard deviation values. It can be noticed that the convergence of the objective function $\mathcal{J}$ for the three optimization algorithms is guaranteed under random initialization values such that $\mathcal{J}$-RSD is approximately around 1% of the objective function mean value $\mathcal{J}$-MEAN as seen in Table 4.5 at various speeds range. In the sense of Paging, TAU, and *Power*, the RSD was almost less than 3% of their associated mean value as shown in Table 4.6. Therefore, the data listed in tables 4.5 and 4.6 prove effectiveness of the three optimization algorithms ABC, PSO and GSA in solving the problem and in particular ABC which recorded the most significant results.

In addition to the aforementioned observations, Table 4.5 also shows that the

Table 4.5: Minimum values of the Five Experiements for the Obective Funtion at Various Speeds Range

| Experiment index | | 1 | 2 | 3 | 4 | 5 | Mean ± STD |
|---|---|---|---|---|---|---|---|
| Speed Range | | 0-8 m/s | | | | | |
| $\mathcal{J}$ | PSO | 199140 | 205820 | 205010 | 200869 | 205777 | 203323 ± 3107 |
| | ABC | 200730 | 194610 | 200350 | 201017 | 196109 | 198563 ± 2981 |
| | GSA | 206120 | 207360 | 205130 | 203749 | 208657 | 206203 ± 1906 |
| Speed Range | | 8-16 m/s | | | | | |
| $\mathcal{J}$ | PSO | 216200 | 213450 | 215930 | 217647 | 212739 | 215193 ± 2039 |
| | ABC | 211250 | 211380 | 212850 | 214280 | 209372 | 211826 ± 1845 |
| | GSA | 217950 | 221770 | 218830 | 221970 | 217062 | 219516 ± 2238 |
| Speed Range | | 16-25 m/s | | | | | |
| $\mathcal{J}$ | PSO | 246930 | 250440 | 241810 | 244939 | 247847 | 246393 ± 3236 |
| | ABC | 239090 | 242860 | 243660 | 244324 | 239416 | 241870 ± 2447 |
| | GSA | 245740 | 245030 | 248280 | 248804 | 243896 | 246350 ± 2114 |
| Speed Range | | 25-33 m/s | | | | | |
| $\mathcal{J}$ | PSO | 296700 | 290450 | 299520 | 298010 | 293102 | 295556 ± 3712 |
| | ABC | 286510 | 289410 | 291510 | 291597 | 286689 | 289143 ± 2482 |
| | GSA | 301690 | 299180 | 294080 | 300770 | 295862 | 298316 ± 3244 |

tracking area update dominates the signaling overhead at all the tested techniques. This is because the value of tracking area update cost $C_t$ is ten folds the value of paging cost $C_\rho$.

Figure 4.5 shows the behavior of the convergence procedure at various speeds range. The results indicate a significant performance difference of the artificial bee colony algorithm among others algorithms. It can noticed from Figure 4.5 that PSO shows faster transient convergence relative to ABC and GSA especially during the first 50 iterations at $v_r = [0, 8]$, $v_r = [8, 16]$, $v_r = [16, 25]$ and $v_r = [25, 33]$. However, as illustrated in Figure

Table 4.6: Average and Standard Deviation Minimum values at last iteration of Convergence for Various Speeds Range

| Speed Range (m/s) | | 0-8 | 8-16 | 16-25 | 25-33 |
|---|---|---|---|---|---|
| Metric | Algorithm | Mean ± STD | | | |
| Paging (Signals) | PSO | 1508 ± 12.5 | 1504 ± 11.2 | 1521 ± 23.8 | 1497 ± 16 |
| | ABC | 1517 ± 23 | 1535 ± 10.5 | 1545 ± 28.4 | 1531 ± 31.6 |
| | GSA | 1503 ± 10 | 1503 ± 18 | 1529 ± 11.1 | 1520 ± 25 |
| TAU (Signals) | PSO | 98589 ± 3026 | 103790 ± 1714 | 118866 ± 3308 | 144436 ± 2826 |
| | ABC | 92190 ± 2586 | 96927 ± 1754 | 109353 ± 2936 | 133503 ± 2391 |
| | GSA | 99958 ± 1972 | 105840 ± 2730 | 117496 ± 2012 | 143680 ± 1254 |
| Power (mW) | PSO | 59.15 ± 1.8 | 124.54 ± 2.05 | 222.87 ± 6.4 | 357.48 ± 7 |
| | ABC | 55.31 ± 1.5 | 116.31 ± 2.1 | 205.03 ± 5.5 | 330.41 ± 6 |
| | GSA | 59.96 ± 1.2 | 127 ± 3.2 | 220.3 ± 3.7 | 355.6 ± 3 |

4.5, ABC shows faster convergence rate to the minimum solution at various speeds range. Figure 4.6 depicts the influence of the tracking area update and paging with regards to the convergence time. Moreover, it is observed that tracking area update and paging have trade-off relation as shown in iteration 250. The objective function adjusts the values of TAU and paging, as the TAU signaling overhead increases the paging decreases due to their trade-off explained in subsection 4.3.2.

However, the artificial bee colony succeeded to give better results for the tracking area update as shown in Figure 4.6 which has the dominant value to influence the objective function as mentioned earlier. Therefore, the artificial bee colony outperforms particle swarm optimization as well as gravitational search algorithm in term of the total signaling overhead. This is further translated in Figure 4.7 that shows the average total signaling overhead at different speed for the applied techniques. It is noticed that ABC technique achieves the minimum values of the total signal overhead as it has been influenced by the lower of triggered TAUs. Whereas the performance of GSA and PSO algorithm is nearly similar as it has been influenced by the increased number of the triggered TAUs that is shown in Figure 4.6.

### 4.5.2 Power Efficiency

In addition to evaluating the total signaling overhead, this subsection aims to evaluate the power consumption of the user equipment's battery. It is estimated that each

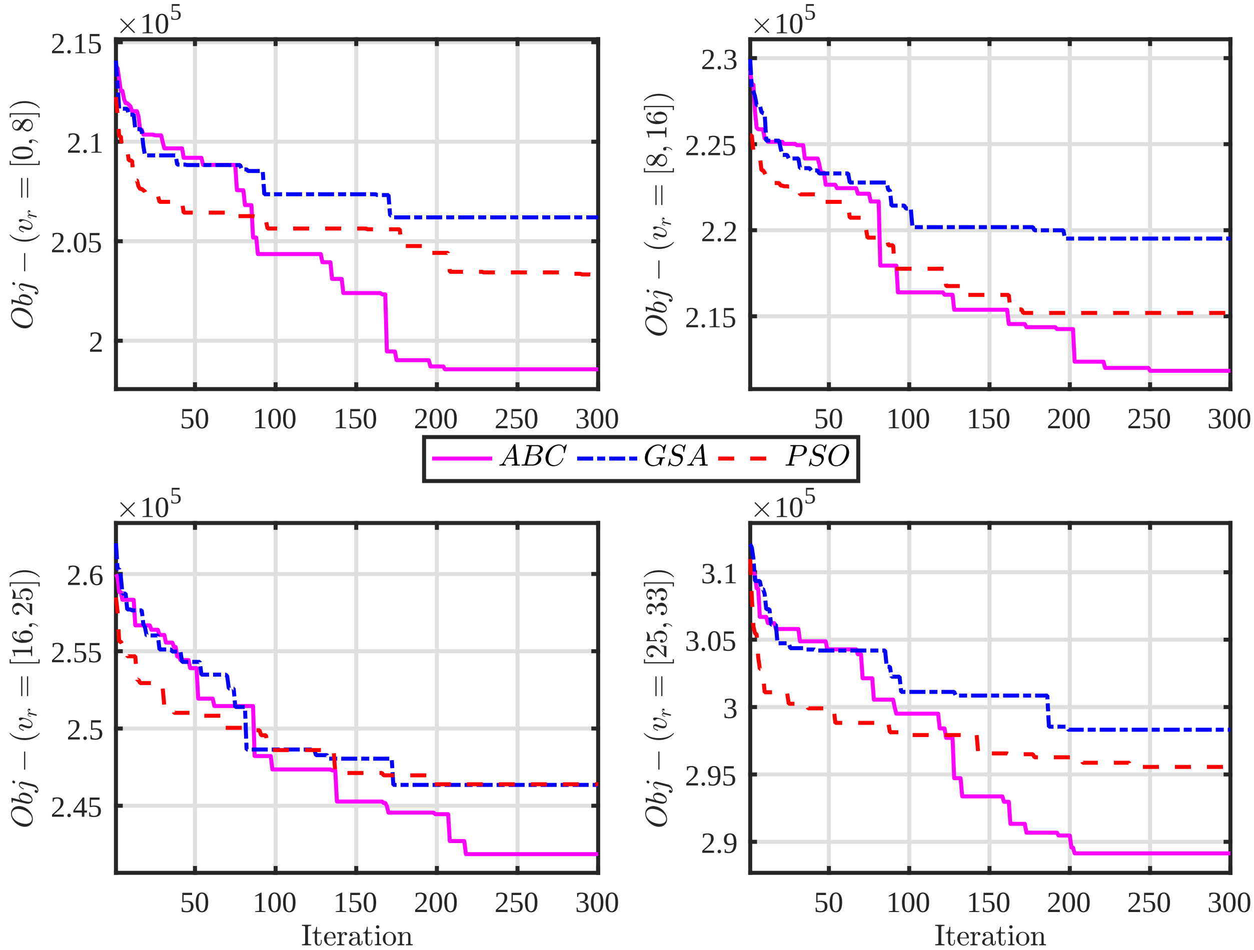


Figure 4.5: Objective function minimization

TAU signal triggered from a regular smart-phone consumes about 10 mW. The values of the total power consumption of each algorithm is shown at the bottom of Table 4.6. The estimated values of the power are taken for each user equipment in hourly basis. In addition, Figure 4.8 shows the converged values of each algorithm. It is clear that the power consumption resulted by using ABC outperforms the other two algorithms in terms of power consumption. This is because the power consumption of the battery is solely dependent on the TAU which is shown to perform better with the ABC algorithm as in Figure 4.6 and Table 4.6. Another observation is that faster speed has a significant impact on the power consumption performance. It is noticed that faster speed can increase the power consumption by approximately six folds.

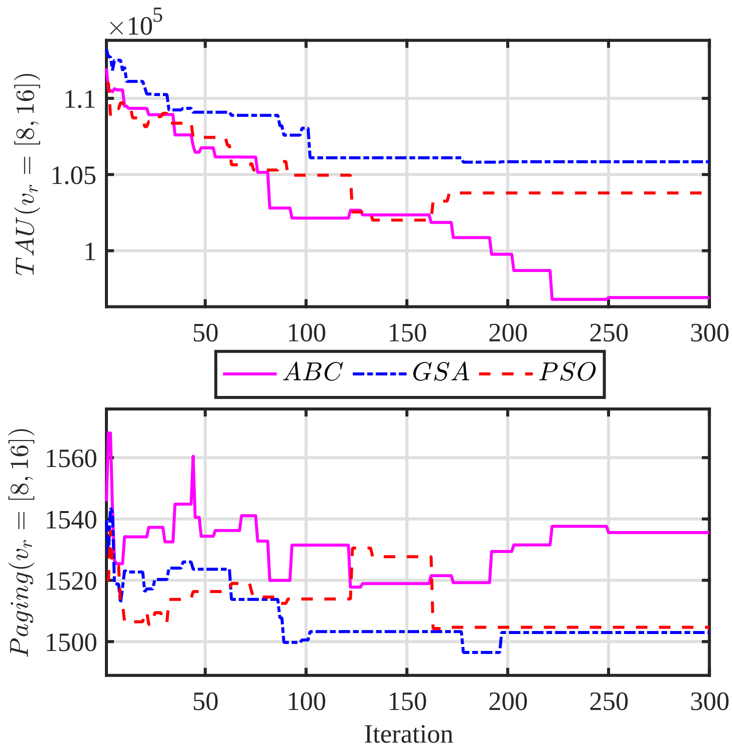


Figure 4.6: TAU and Paging Average Signaling Overhead

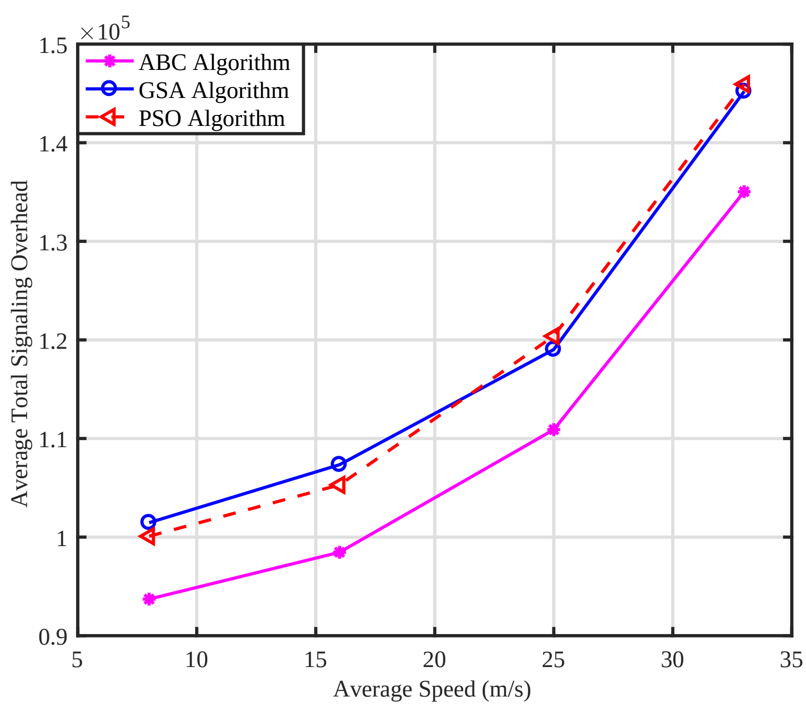


Figure 4.7: Average Total Signaling Overhead

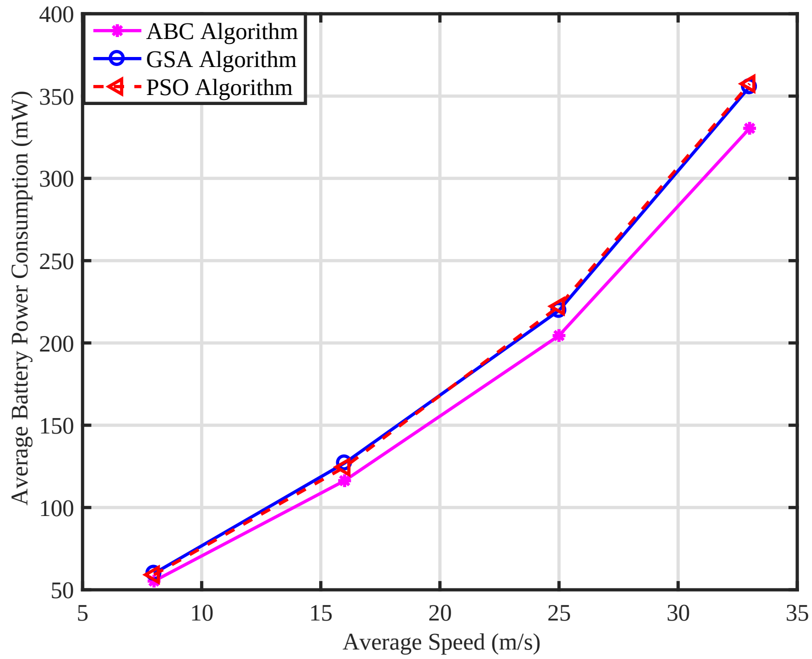


Figure 4.8: Average Battery Power Consumption (mW)

## 4.6 Conclusion

Cellular networks have been one of the most vital technology enablers allowing a large spread of network connectivity. In recent years, LTE and LTE advance are considered the fastest wireless networks. However, the rapid increase of new devices and various applications pose various performance challenges, especially in terms of signaling overhead. Consequently, this can raise a crucial concern related to the availability of the system. One of the important causes of signaling overhead is location management that allocates the idle users and tracks their movement within the system. Tracking area update and paging are triggered whenever the idle user moves from one cell to another that is not within the same tracking area list and to page an idle user respectively. In this study, tracking area update and paging have been investigated with the aim to minimize the total signaling overhead. A centralized TAL-to-MME scheme is used as it has been explored previously [22] to be the most suitable design for minimizing the signaling overhead. Three evolutionary algorithms have been used that belong to different heuristic search families namely, Particle Swarm Optimization, Artificial Bee Colony, and Gravitational Search Algorithm. The three algorithms' convergence has been tested to obtain the minimum value. It is found that the minimum values of the three algorithms are guaranteed. Five experiments have been conducted under different random initialization values and it is found that the relative standard deviation for the objective function is around 1%, whereas it was around 3% for the paging, tracking area update, and power. Hence, the three applied optimization algorithms have proven to be efficient for solving the problem. Furthermore, ABC has recorded the best convergence results among others algorithms.

# Chapter 5
# Power-Aware Optimized RRH to BBU Allocation in C-RAN

## 5.1 Introduction

Since the last decade, wireless networks have gained tremendous attention allowing new technologies to rise and favoring a vast increase in the number of hand-held devices. On the other hand, the proliferation of new applications and the exponential growth rate of devices per user have affected the performance of wireless technology in various ways. First, there has been an impact on performance caused by the tremendous volume of data traffic along with the total signaling overhead because of the rapid movement of a large number of users [36]. These problems can degrade the level of service, especially what is known as the Quality of Experience (QoE) of the users. Part of the impact can be seen as loss of availability and extensive delay. This indeed triggers a demand for new concepts and technologies. Several researchers have striven to introduce better and more effective solutions that favor availability, reliability, cost efficiency, and Quality of service.

In order to face the aforementioned challenges, Cloud Radio Access Network (C-RAN) has been introduced as a new paradigm that has succeeded in bringing forth a new era to the world of wireless communications. C-RAN was first introduced by China Mobile, one of the key wireless operators in China. The advantages of cloud computing have influenced researchers in both academia and industry to integrate the cloud paradigm into different applications. Featuring a simple yet clever solution that splits up conventional base stations into two independent entities, namely a Remote Radio Head (RRH) and a Baseband Unit (BBU), connected through an optical high-speed transport network. Cloud computing enables real-time centralized processing through a virtualized BBU pool. In addition, the cloud architecture provides several advantages for C-RAN in numerous aspects, such as reducing the total CAPEX and OPEX as well as providing flexibility by distributing the capacity of the system.

The separation of the computational resources from the RRH has resulted in a significant reduction of power and has increased spectral efficiency. Moreover, the RRHs are deployed as small cells that can be densely distributed in a way that causes minimum interference. Thus, the distance between the user equipment (UE) and RRHs is minimized, allowing for a more stabilized throughput gain. In addition, the implementation of BBU pools help to alleviate the energy cost of transmission and reception between the RRH and the BBU.

However, the efficiency of C-RAN is heavily dependent on the processing resources available in the BBU pool. In other words, there is a tight correlation between the efficiency of the mapping of computing resources from BBU to RRH and the overall performance of the system. The dense deployment of small cell imposes the necessity of distributing and allocating the resources for RRHs in the BBU pool intelligently. Thus, resource scheduling between RRH and BBU is essential for efficient and reliable C-RAN implementation. In order to quantify these resources, the system calculates the computing requirements of each RRH and accordingly distributes the resources available within the BBU pools.

On the other hand, the computing requirements of each RRH are related to the scheduling of the physical resources to the users such that the Quality of Service (QoS) is satisfied. Because of this, the system has two levels of scheduling, each of which carries the same importance. The system first ensures user satisfaction by suitably distributing the physical resources and then schedules the baseband processing requirements simultaneously to all the BBU pools. Therefore, it is vital to optimize the scheduling process at each level. Optimizing computing resources in cloud systems is important because of the increased demand caused by intensive applications. In addition, network operators seek to minimize the cost of expanding resources while also having adequate computing resources available for instant processing.

In this work, a formulated optimization model depicting the two levels of scheduling is presented. To the best of our knowledge, this study is the first to propose a complete optimization model formulating a strategy for both ends. In the level comprised between cells and users, resources are distributed among users, which have different QoS requirements. As a consequence, the system has to optimize resource allocation accordingly while maintaining other aspects such as availability of physical resources, satisfying the QoS, and continuity of service. In the RRH and BBU level, computing requirements need to be processed instantly in the available BBU pool available while maintaining power consumption and optimizing computing resources. This model is an NP-HARD problem because of the two levels resource assignments. The problem of finding optimal assignment is known to be intractable to be solved in a polynomial time [17, 18]. Therefore, the model is simplified into two sub-problems using the proposed decomposition model. In addition, two heuristic solutions of lower complexity are proposed to solve both problems.

The rest of this chapter is organized as follows. Section 5.2 presents literature within the field of mobility management techniques. Section 5.3 presents a detailed description of the system model and discusses presents the formulation of the problem along with its complexity analysis. The simulation parameters and results are discussed in Section 5.4. Finally, Section 5.5 concludes the chapter.

## 5.2 Related Work

C-RAN is an emerging technology that is vastly growing at a fast pace. As a result, C-RAN has succeeded in gaining the attention of both the academia and the industry. Various studies have been made to tackle a number of aspects in order to overcome the challenges that might be encountered. These studies can be categorized into three levels, where each level is focused on the main elements of C-RAN: RRH, BBU, and fronthaul. In this section, several studies are highlighted in order to shed light on the latest advancements in C-RAN optimization.

In [54], authors discussed the computing capability needed to run C-RAN. The authors managed to model the relative computing resource consumption that corresponds to the users' end by characterizing the relationship between resource consumption in the BBU and the signal transmitted to the users. The authors concluded that the computing resources consumed by the BBU increase significantly in a non-linear fashion. They also suggested that user distribution has a critical impact on BBU computing resource consumption.

Authors in [55] attempted to enhance wireless access capability by using Virtual Base Stations (VBSs) which help in reducing interference. Each virtual base station is a cluster of multiple base stations, which can be freely used by the users in the system. Despite the novelty of the idea, numerous strategies might be required to realize "no-edge" wireless networks, such as using different time slots and sub-bands. Moreover, an advanced power control mechanism is needed for guaranteeing the QoS delivered to each user.

Another dynamic clustering algorithm was presented in [56], which used a greedy multi-objective optimization technique. The authors sought to improve the energy efficiency of the RRHs by using a clustering technique, while also increasing the joint capacity of RRHs in the system. A linear model along with a scalarization method were posited in order to solve the non-linear optimization problem. The designed model has been implemented in the downlink.

A different technique was proposed in [57], where the authors designed a greedy algorithm for the purpose of minimizing power consumption of both the RRHs and the transport network. The authors focused on the joint selection of RRHs by enabling an on and off mechanism that resulted in minimizing the power consumption of the transport network. Moreover, the ability of turning off RRHs was achieved with the help of two group sparse beamforming methods, namely the bisection and iterative methods.

In [58], a self-organizing C-RAN was presented for the sake of improving the QoS at the end points. The idea of the study was to reduce the number of blocked cells in the system in such a way that increases the QoS of the users. The self-organization mechanism aimed to dynamically adjust the mapping of resources between BBUs and RRHs with the help of a third-party server. The server, called "Host Manager", was located inside the BBU pool, which optimizes the load conditions, and chooses the best configurations using genetic algorithms. However, the messages and the complexity of the added server is unknown, and can be exponential.

In [59], the authors proposed a load-aware method that minimizes the number of active BBUs required to process RRH computational resource demands. The authors introduced what they called a lightweight load-aware algorithm that dynamically optimizes the BBU pooling gain. The mapping of RRHs to BBUs is based on many-to-one mapping. The authors claim that their approach delivers an almost optimal performance, by using the well-known First Fit Decreasing (FFD) algorithm.

An interference-aware resource allocation and consolidation (CoRAC) TDD C-RAN is proposed in [60, 61]. The study presents a dynamic method for re-configuring and mapping the resources between RRHs and BBUs based on the utilization status of the BBUs and the cross-sub-frame Co-Channel Interference (CCI). This approach relies on virtually clustering the cells, making the downlink (DL) and uplink (UL) configurations dependent on the performance of a group of cells. Thus, performance gain depends on the cross relation of the DL/UP physical resources and the RRH-to-BBU mapping. Although this cross multi-objective optimization is desirable, the physical re-configuration and adaptation of the uplink and downlink may highly increase the complexity of the system.

In [62], the authors proposed two-stage iterative heuristic algorithm aims to minimize the network power consumption of C-RAN. The proposed solution focuses on the hybrid backhaul that constitutes of a wireless and wireline connections. The first stage of the solution solves the network power consumption by implementing a joint BS selection and beanforming optimization problem. The second stage is for solving the transmit power consumption of the wireless links. The study, however, has not considered the power consumption of either the BBU pool or the cloud physical resources.

Energy efficient resource assignment allocation in Heterogeneous C-RAN was studied in [63, 64] . Authors in [63] attempted to maximize the energy efficiency in the OFDMA-based H-CRAN that is correlated to RB and power assignments. An enhanced Soft Fractional Frequency Reuse (S-FFR) scheme was proposed in order to ameliorate the performance of users that are located in the cell center zone. In [64], the authors investigated the energy efficient of the Control Data Separation Architecture (CDSA)-based H-CRAN. The authors introduced a model for alleviating the power consumption for the CDSA-based H-CRAN networks. The model is constrained by the average minimum data rate and the limited capacity of the fronthaul. Both studies [63, 64] have not depicted the impact of the proposed model on the BBU pool.

## 5.3 System model

### 5.3.1 Problem definition

This work attempts to minimize the power consumption in computer resources that results from processing the cellular network baseband computational requirements. In order to achieve the aforementioned goal, the number of baseband units (BBUs) used to process the baseband computational requirements need to be minimized. The BBUs' use is related to the users' requirements. Therefore, the model satisfies the Quality of Service requirements of the users by maximizing the aggregated data rate while minimizing the power consumption of the BBUs. Thus, the model achieves a high system performance by delivering an enhanced user Quality of Experience while reducing the power consumption that occurs in the processing units.

### 5.3.2 Preliminaries

In this work, the downlink transmission of C-RAN is considered. Let $|M|$ be the total number of BBUs; each BBU $j \in M$ serves a number of RRHs by offering their processing resources, which correspond to three main functions related to baseband processing, namely coding, modulation, and FFT. The system has a total of $|N|$ cells/RRHs, and each RRH is denoted as $i$ where $i \in N$. According to [65] the computing resources requirements for modulation and coding can be approximated as a linear function of the modulation and coding scheme (MCS). The MCS is therefore estimated by the transmission data rate associated to each user. In contrast, the FFT function has constant computational requirements. The computing resource requirements for the baseband processing for each user can be written as follows [54]:

$$L_k^b = \theta \log(1 + SINR) + L_B \tag{5.1}$$

where $\theta$ is the experimental parameter and $L_B$ is the constant complexity of the FFT function. For each cell $i$ the computing resource requirements for coding, modulation and FFT 50%, 10% and 40%, respectively [66].

### 5.3.3 Channel Model

Both the distance-dependent macroscopic path loss ($MPL$) and the shadow fading path loss components are considered. Equation (5.2) models the macroscopic path loss between the RRH and a user at distance $d$ in an urban environment [67]:

$$\begin{aligned} MPL_{dB}(d) = {} & 40(1 - 4 \times 10^{-3} h_b) \log_{10}(d/1000) \\ & -18 \log_{10}(h_b) + 21 \log_{10}(f_c) + 80 \end{aligned} \tag{5.2}$$

where $h_b$ represents the base station antenna height (in meters) and $f_c$ represents the carrier frequency in MHz.

On the other hand, it is assumed that the shadow fading path loss component can be represented as a Gaussian random variable with standard deviation $\sigma$, expressed in dB. Hence, the total path loss between the RRH and the user is as follows: [67]

$$PL_{dB,(RRH,k)}(d) = MPL_{dB}(d) + \log_{10}(X_k) \tag{5.3}$$

where $X_k$ represents the log-normal shadow fading path loss of user $k$. Consequently, the linear gain between the RRH and user $k$ is as follows: [67]

$$G_{(RRH,k)} = 10^{-PL_{dB,(RRH,k)}/10} \tag{5.4}$$

### 5.3.4 Design Hypothesis

There are two essential components of the C-RAN architecture that are considered, namely the RRHs and the pool of BBUs where all the baseband processing is performed. We can say that a number of virtual BBUs are coupled and mounted onto several physical machines located in the data center. Therefore, each physical machine can represent a centralized pool containing a number of BBUs. Moreover, a single BBU can have either a one-to-one or a one-to-many relationship with the RRHs.

There are a finite number of BBU entities mapped into a BBU pool, with each BBU entity having an equivalent computing processing capacity represented as Million Operations Per Time Slot (MOPTS) and a finite number of cells serving the user equipment. The processing capacity available at the pool of BBUs is greater than the total computer resources required for all the cells in the system. Figure 5.1 depicts the architecture of the system, and shows multiple RRHs colored differently. These RRHs are served by the BBUs available in their coverage area. Moreover, the same coloring scheme is applied at the BBU end, which indicates that the BBUs serve a number of RRHs within their coverage range. It should be noted that RRHs have variable processing requirements, which are related to the users' requirements during each time slot. There are a total of $|K|$ users in the system, each of which is denoted as $k$, where $k \in K$. RRHs are connected to the BBU pool via an ideal fronthaul that has enough capacity resources. Table 5.1 explains the different notations that are used in the system model.

### 5.3.5 Problem Formulation

The problem expressed in (5.5) was formulated to minimize the number of BBUs used to process the computational requirements while maximizing the users' data rate. Equation (5.5a) presents $\alpha$ as the number of bins to be minimized, as described in the cost function (5.5b), and presents $\beta$ as the data rate to be maximized, as given in the second cost function (5.5c). The problem is subject to various constraints (5.5d, 5.5e, 5.5f, and 5.5g) to ensure that the limited computational resources that each bin can have is not violated, that each RRH is assigned to a BBU, and that the minimum rate requirement for each user is met.

- **Objective function**

$$\min \Big[\alpha - \beta\Big] \tag{5.5a}$$

- **Cost functions**

$$\alpha = \sum_{j}^{|M|} \sum_{i}^{|N|} \sum_{k}^{|K|} O_k^i . L_k^b . B_j \tag{5.5b}$$

$$\beta = \sum_{k=1}^{|K|} \sum_{rb=1}^{|RB|} \left[ B.\, x_k^{rb} . \log_2 \left( 1 + \frac{P_{(i,k)}^k . G_{(i,k)}^{rb}}{N_0 . B} \right) \right] \tag{5.5c}$$

**Subject to**

$$\sum_{k=1}^{|K|} x_k^{rb} = 1;\ \forall\, rb \in RB \tag{5.5d}$$

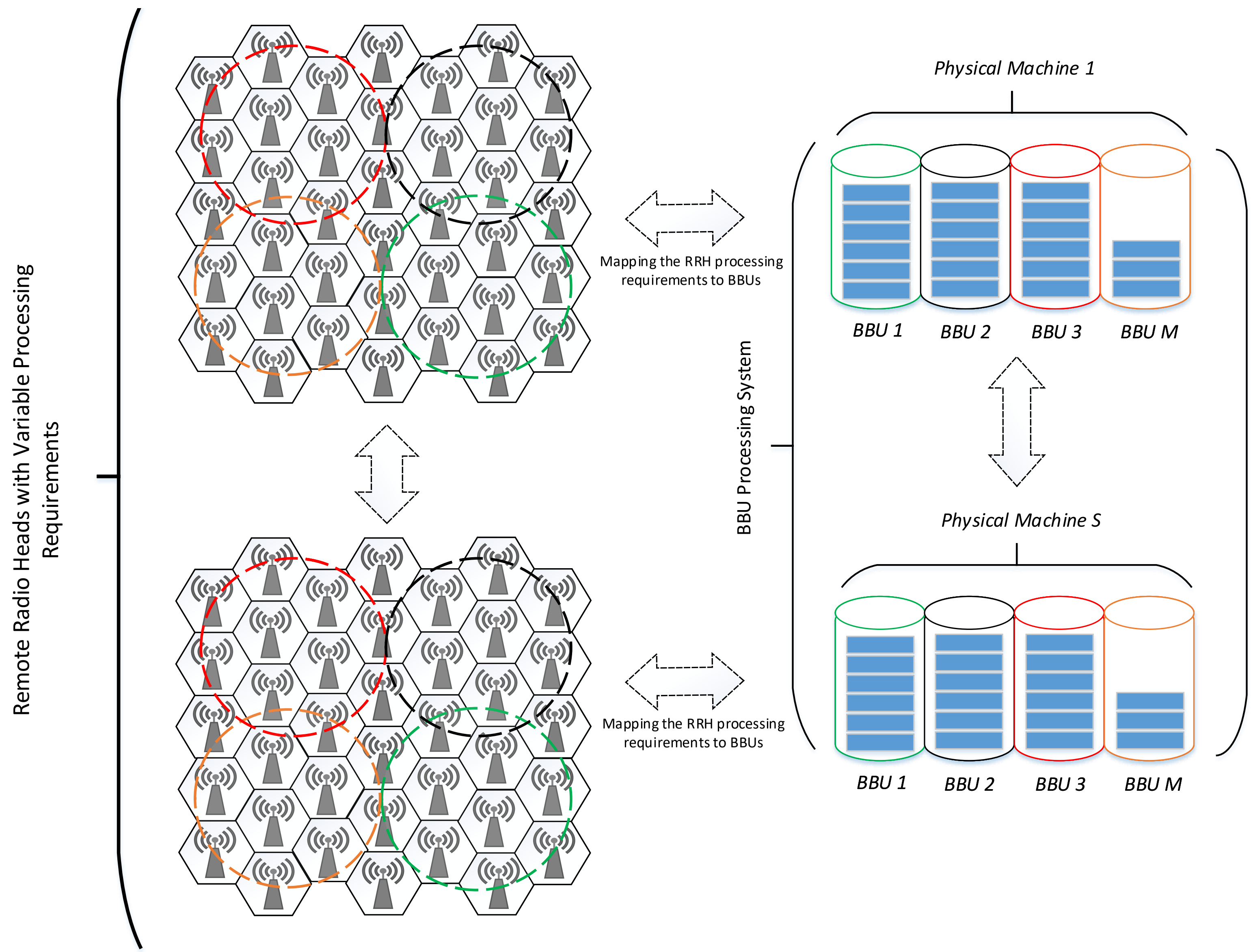


Figure 5.1: RRH to BBU mapping architecture

Table 5.1: Table of Notations

| | | |
|---|---|---|
| $P_{pool}$ | : | Power consumption of the BBUs pool. |
| $P_j^{tot}$ | : | Total power consumption of an individual BBU $j$. |
| $P_i$ | : | Power consumption of RRH/Cell $i$. |
| $P^{TOT}$ | : | Total power consumption. |
| $P_{st}^{pool}$ | : | Static power consumption of an individual pool $s$. |
| $P_{st}^{j}$ | : | Static power of each BBU obtained by well-known procedures. |
| $L^j$ | : | Available computing resources in each BBU $j$. |
| $L_i^j$ | : | Total computing resources given by BBU $j$ to cell $i$. |
| $L_J^{tot}$ | : | Total computing resources used by BBU $j$. |
| $\|N\|$ | : | Total number of RRHs/Cells. |
| $\|M\|$ | : | Total number of BBUs. |
| $\|RB\|$ | : | Total number of RBs. |
| $\|K\|$ | : | Total number of users. |
| $B$ | : | RB bandwidth (typically 180 KHz). |
| $P_{(i,k)}$ | : | Constant transmission power from RRH $i$ to user $k$. |
| $G_{(i,k)}^{rb}$ | : | Channel condition between RRH $i$ and user $k$ on RB $rb$ |
| $N_0$ | : | Noise |
| $r_{k,th}$ | : | Minimum data rate requirement of user $k$. |
| $O_k^i$ | = | $\begin{cases} 1, & \text{if user } k \text{ is served by cell } i, \\ 0, & \text{otherwise} \end{cases}$ |
| | | Decision Variables |
| $B_j$ | = | $\begin{cases} 1, & \text{if BBU } j \text{ is used,} \\ 0, & \text{otherwise} \end{cases}$ |
| $X_i^j$ | = | $\begin{cases} 1, & \text{if cell } i \text{ uses BBU } j, \\ 0, & \text{otherwise} \end{cases}$ |
| $x_k^{rb}$ | = | $\begin{cases} 1, & \text{if user } k \text{ is allocated RB } rb, \\ 0, & \text{otherwise} \end{cases}$ |

$$\sum_{i}^{|N|} L_i.X_i^j \leqslant B_j.L^j; \quad \forall j \in M \tag{5.5e}$$

$$\sum_{i}^{|N|} X_i^j = 1; \quad \forall j \in M \tag{5.5f}$$

$$\sum_{rb=1}^{|RB|} B.\, x_k^{rb}.\log_2\left(1 + \frac{P_{(i,k)}.G_{(i,k)}^{rb}}{N_0.B}\right) \geq r_{k,th};\ \forall\, k \in K \tag{5.5g}$$

### 5.3.6 Decomposition Model

The problem stated in (5.5) is a binary integer programming (BIP) problem. Consequently, finding the global optimal solution can be computationally expensive due to the large size of the search space for both the resource-block-to-user-allocation as well as the RRH-to-BBU allocation. Therefore, different techniques can be used to solve the problem. A decomposition model that divides the problem into a two-stage resource allocation problem is proposed and can be described as follows:

**Sub-Problem 1 Definition**. *In a given system that has $N$ cells that serve $K$ UEs, find the optimum allocation of physical resources in each cell in order to satisfy the QoS requirements of the users in terms of data rate. In other words, the system strives to maximize the data rate for the users under the minimum rate requirement constraint. The baseband processing requirements that are needed to maximize the data rate for each user are aggregated and ready to be sent to the next stage of allocation. Figure 5.2 shows that each user has three different tasks involving baseband processing requirements.*

**Sub-Problem 2 Definition**. *For a given set of computational requirements $Li$ expressed in MOPTS which arise from each cell $i \in N$, and a set of BBUs $M$ each of which has $L^j$ processing resources pooled into a set of physical clusters, find the optimal allocation of $Li$ to be packed into an optimized number of BBU processing units, $B_j$, so that total power consumption is minimized. The assignment of cell resources to the BBU is denoted as a binary variable, $X_i^j$.*

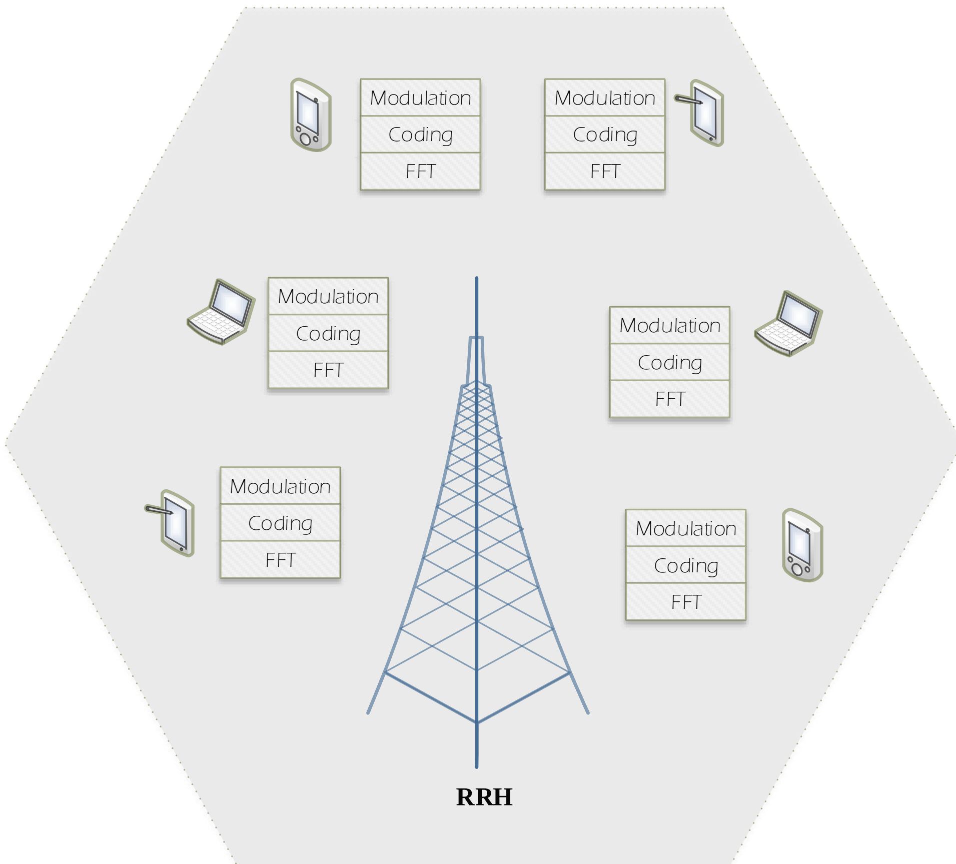


Figure 5.2: Users to RRHs baseband processing requirements.

#### 5.3.6.1 Sub-Problem 1 formulation

This part of the problem is to ensure that the users' QoS requirement is satisfied by maximizing the aggregated data rate for the users. The following equations formulate the objective function and the constraints for yielding maximum aggregated data rate for the users allocated in each cell:

- **Objective function**

$$\max \beta = \left[ \sum_{k=1}^{|K|} \sum_{rb=1}^{|RB|} B.\, x_k^{rb}.\log_2 \left( 1 + \frac{P_{(i,k)}.G_{(i,k)}^{rb}}{N_0.B} \right) \right] \tag{5.6a}$$

**Subject to**

$$\sum_{k=1}^{|K|} x_k^{rb} = 1;\ \forall\, rb \in RB \tag{5.6b}$$

$$\sum_{rb=1}^{|RB|} \left[ B.\, x_k^{rb}.\log_2 \left( 1 + \frac{P_{(i,k)}.G_{(i,k)}^{rb}}{N_0.B} \right) \right] \geq r_{k,th};\quad \forall\, k \in K \tag{5.6c}$$

$$x_k^{rb} \in \{0, 1\}; \forall rb \in RB, \forall\, k \in K \tag{5.6d}$$

where $x_k^{rb}$ is a binary variable indicating whether user $k$ is allocated RB $rb$, as shown in constraint (5.6d), and $B$ is the bandwidth of an RB. Constraint (5.6b) states that an RB can only be assigned to one cell user. Constraint (5.6c) is the minimum rate requirement for each cell user.

The complexity of the above sub-problem for each cell is of order $O(2^{|RB.K|})$, where $|RB|$ is the total number of RBs within the cell and $|K|$ is the total number of users. As a result the total complexity for all the cells within the system is $O(|N|.2^{|RB.K|})$, where $|N|$ is the total number of cells.

#### 5.3.6.2 Sub-Problem 2 formulation

This part is conventionally formulated as a bin packing problem that represents each BBU as a bin with finite computing resources, expressed in MOPTS. On the other hand, the cell traffic loads are defined as items that need to be packed into BBU bins.

In this context, we intend to model each aspect of the system that is related to the baseband computational resources required by the users. The baseband processing requirements of users are aggregated at the cell level during every time slot and are mapped to the BBU pool. As discussed earlier, the objective is to minimize the number of BBU units so that the total power consumed by the system is minimized. Therefore, the problem expressed in (5.7) aims to minimize the total number of BBUs $B_j$ serving the cells (5.7a). Equation (5.7b) aggregates the baseband computing requirement of each user $k$ served by each cell $i$. The variable $O_k^i$ is a binary indicator showing whether user $k$ is served by cell $i$, and $L_k^b$ are the baseband processing requirements of user $k$. Equation (5.7c) calculates the total computational resource consumption of each BBU $j$ in the system, obtained from the number of cells assigned to it. Equation (5.7) subject to multiple constraints. Constraint (5.7d) limits the computational resources required for the cells served by a specific BBU $j$ to its maximum computational resources $L^j$. Constraint (5.7e) ensures that each cell $i$ is assigned to one BBU $j$, while more than one cell can be assigned to a single BBU $j$. Lastly, constraints (5.7f) and (5.7g) limit the value of binary variables $X_i^j$ and $B_j$ to be either 0 or 1.

- **Objective function**

$$\min \alpha = \sum_{j}^{|M|} \sum_{i}^{|N|} B_j \tag{5.7a}$$

- **Cost functions**

$$L_i = \sum_{k}^{|K|} O_k^i . L_k^b; \quad \forall i \in N \tag{5.7b}$$

$$L_j^{tot} = \sum_{i}^{|N|} L_i . X_i^j; \quad \forall j \in M \tag{5.7c}$$

**Subject to**

$$\sum_{i}^{|N|} L_i . X_i^j \leqslant B_j . L^j; \quad \forall j \in M \tag{5.7d}$$

$$\sum_{i}^{|N|} X_i^j = 1; \quad \forall j \in M \tag{5.7e}$$

$$X_i^j \in 0, 1; \quad \forall i \in N \tag{5.7f}$$

$$B_j \in 0, 1; \quad \forall j \in M \tag{5.7g}$$

The complexity of the above sub-problem is of order $O(2^{|N|}.N)$, where $|N|$ is the total number of cells within the system. This is because the maximum possible number of user bins is equal to the number of cells to be packed. Furthermore, there are $2^{|N|}$ possible packing combinations of the cells within each bin.

#### 5.3.6.3 Decomposition Model Mechanism

As the name of the model suggests, the original problem is divided into two sub-problems which can be sequentially solved. Algorithm (4) presents a way of solving the two sub-problems. Line 1 shows the input of the problem, which is the total resource blocks, the number of cells, the available BBUs, and the total number of users served by each cell. Line 2 shows the output, which are the aggregated baseband computing requirements each cell demands and the allocated computing requirements for all cells that are packed into the BBUs. Lines 2 to 7 are for solving the first sub-problem for each user $k$ served by cell $i$. The goal is to determine the total baseband resources required at each cell in the system. Line 8 is to calculate the processing requirements of each cell in terms of MOPTS. Lines 10 to 15 are for solving the second sub-problem, which is to process the required baseband resources for each cell into the BBU pool in accordance with their maximum capacity. The output of this algorithm is the assignment of cells to BBUs, the BBUs used, and the total utilization of each BBU.

### 5.3.7 Heuristic Algorithm

In this section, a heuristic algorithm is developed in order to solve a problem with a lower complexity than the decomposition model.The heuristic algorithm follows the same concept of the decomposition model by dividing the problem into two sub-problems, namely a resource block allocation problem and a bin packing problem. The first sub-problem of the heuristic algorithm seeks to first satisfy the minimum data rate requirement by efficiently allocating the resource blocks to users in a sequential manner. This insures that the proposed solution delivers the Quality of Service (QoS) requirements to the users. The second step of the first sub-problem is to allocate the remaining physical resource blocks in such a manner that maximizes the aggregated data rate for the users.

**Algorithm 4** Decomposition Mechanism

**Input:** $rb = \{1, 2, ..., RB\}$: Total number of Resource Blocks
  $i = \{1, 2, ..., N\}$: Total number of cells
  $j = \{1, 2, ..., M\}$: Total number of BBUs
  $k = \{1, 2, ..., K\}$: Total number of users served by each cell
**Output:** $L_i$: The aggregated baseband computing requirements of each cell and the allocation of computing resources for all cells associated to BBUs.
**for** $i \in N$ **do**
  **for** $k \in K$ **do**
    **Solve the first sub-problem** (5.6)
    **calculate** $L_k^b$ from Equation (5.1)
  **end for**
  **calculate** $L_i = \sum_{k}^{|K|} O_k^i . L_k^b$
**end for**
**for** $i \in N$ **do**
  **for** $j \in M$ **do**
    **Solve the second sub-problem** (5.7)
    **Output** $X_i^j$, $B_j$, and $L_j^{tot}$
  **end for**
**end for**

The second sub-problem is to pack the aggregated baseband computational requirements of each cell into the BBU pool in such a way that the power consumption of the physical resources is minimized. A first fit decreasing bin packing algorithm is used in this part to minimize the number of BBUs used. The computational requirements of the cells are modeled as items to be packed into the BBUs, which are modeled as bins.

Algorithm (5) presents the proposed heuristic algorithm along with its mechanism. In the first part (lines 3 to 12), each user's minimum rate constraint is satisfied by sequentially allocating resources to the users. To do so, RBs with the best channel gains $G_{(i,k)}^{rb}$ are allocated for each user until the $r_{k,th}$ threshold is reached, which is calculated based on if the rate expression in equation (5.6a) satisfies the minimum rate. In the second part (lines 13 to 17), the remaining RBs are allocated according to the channel gains in such a manner that would maximize the aggregated user rate. In line 18, the algorithm then calculates the baseband processing requirements of each user $L_k^b$. The requirements are aggregated at the cell level, as shown in line 19. The second sub-problem is solving the allocation of the computational requirements of each cell $Li$ to the BBUs $M$ at the backhaul. The computational requirements are modeled as items to be packed in bins that represent the BBUs. The bin packing algorithm that is used is the first fit decreasing bin packing algorithm written in lines 24 to 32. The FFD sorts the computational requirements of each cell in a decreasing order. After that, each cell $i \in N$ is packed into an available BBU $j \in M$ in the system. The output of the algorithm is the set of BBUs that are specified to serve cells $B_j$, the assignment of cells to BBUs $X_i^j$, and the total computational resource consumption of each BBU $L_j^{tot}$.

### 5.3.8 Complexity Analysis

This subsection discusses the complexity of each of the proposed algorithms.

#### 5.3.8.1 Decomposition Model

The complexity of the decomposition algorithm sub-problem 1 for each cell is of order $O(2^{|RB.K|})$, where $|RB|$ is the total number of RBs within the cell and $|K|$ is the total number of users. As a result the total complexity for all the cells within the system is $O(|N|.2^{|RB.K|})$, where $|N|$ is the total number of cells.
Whereas the complexity of the decomposition algorithm sub-problem 2 is of order $O(|N|.2^{|N|})$, where $|N|$ is the total number of cells within the system. This is because the maximum possible number of user bins is equal to the number of cells to be packed. Furthermore, there are $2^{|N|}$ possible packing combinations of the cells within each bin. Hence, the total complexity of the decomposition model is $O(|N|.(2^{|RB.K|} + 2^{|N|}))$.

#### 5.3.8.2 Heuristic Algorithm

The order of complexity of the heuristic algorithm targeting the first sub-problem is $O(|K||RB|)$, where $|RB|$ is the total number of RBs available and $|K|$ is the total number of users in the system. In addition, the complexity of the heuristic algorithm for solving the second sub-problem is of order $O(|N| \log |N|)$, where $|N|$ is the total number of cells. Therefore, the total complexity of the heuristic algorithm is $O(|K||RB| + |N| \log |N|)$.

**Algorithm 5** Heuristic Algorithm/Mechanism

**Sup-Problem 1**
**Input:** $rb = \{1, 2, ..., RB\}$: Total number of Resource Blocks
$i = \{1, 2, ..., N\}$: Total number of cells
$k = \{1, 2, ..., K\}$: Total number of users served by each cell
**Output:** $L_i$: The aggregated baseband computing requirements of each cell.
**for** $i \in N$ **do**
  **for** $k \in K$ **do**
    **while** $r_k < r_{k,th}$ **do**
      **find** $g_{k,rb} = \max_{rb \in RB} \{G^{rb}_{(i,k)}\}$
      **set** $x^{rb}_k = 1$
      **calculate** $r_k$
      **update** $RB = RB \setminus rb$
    **end while**
  **end for**
  **for** $rb \in RB$ **do**
    **find** $g_{k,rb} = \max_{rb \in RB} \{G^{rb}_{(i,k)}\}$
    **set** $x^{rb}_k = 1$
    **update** $RB = RB \setminus rb$
  **end for**
  **calculate** $L^b_k$ from Equation (5.1)
  **calculate** $L_i = \sum_{k}^{|K|} O^i_k . L^b_k$
**end for**
**Sup-Problem 2**
**Input:** $j = \{1, 2, ..., M\}$: Total number of BBUs
$Li$ where $i \in N$: Computing requirements of cell $i$
**Output:** $X^j_i$, $B_j$, and $L^{tot}_j$
**Sort** $L_i$ **of all cells in decreasing order**
**for** $i \in N$ **do**
  **for** $j \in M$ **do**
    **if** $L_i \in i$ fits in BBU $j$ **then**
      Pack $L_i$ in BBU $j$.
      Break the loop and pack the next object $L_i$.
    **end if**
  **end for**
**end for**

### 5.3.9 Generalized C-RAN Energy Consumption (GCEC) Model

In this context, the energy model presented in Equation (5.8) is based on the previous model posited in [68], which was inherited from the well-known EARTH model [69]. However, we have generalized the model so that we can analyze the power consumption of our system and any other model related to C-RAN. In the original model, the authors calculated the power of the BBU pool by adding the power of a group of RRHs to the power of the single BBU that served them. This assumption can redundantly add more power for the same RRH if the system allows more than one BBU to serve a single RRH. In contrast, the generalized model calculates the total power consumption by concatenating the added power consumed by all the BBU pools and all the RRHs simultaneously in each time slot. This allows for either using a single or multiple BBUs to serve each RRH in the system.

Equation (5.8b) evaluates the power consumption of each pool of BBUs by adding the total power consumed by each BBU $j$ belonging to the pool and the static physical power of the hardware that is hosting the pool. The power consumption of each BBU is calculated in (5.8c) by adding the total resources used in the BBU (noted as $L_J^{tot}$ and multiplied by the weight factor of power $\omega$) and the static power consumption of the well-known procedures related to BBUs. Moreover, equation (5.8d) shows that the power consumed by each RRH relies on the number of resource blocks that are allocated to each user, the power amplifier efficiency $PA$, the radio frequency power consumption $RF$, and the number of transceiver antennas $N_{TRX}$. The antenna output power is calculated in equation (5.8e), which was obtained from the EARTH model [32].

- **GCEC Model**

$$P^{TOT} = \sum_{s}^{|S|} P_{pool} + \sum_{i}^{|N|} P_i \tag{5.8a}$$

$$P_{pool} = \sum_{j}^{|M|} P_j^{tot} + P_{st}^{pool} \tag{5.8b}$$

$$P_j^{tot} = \omega.L_J^{tot} + P_{st}^{j}; \quad \forall j \in M \tag{5.8c}$$

$$P_i = \frac{P_{out}}{PA} + P_{RF}.N_{TRX} \tag{5.8d}$$

$$P_{out} = m.|RB| + n \tag{5.8e}$$

## 5.4 Performance Evaluation and analysis

### 5.4.1 Simulation Parameters:

A system simulation using MATLAB was performed on an Intel Core i7-4770 CPU by adopting a realistic LTE system model. The system was tested under three different configuration options (150, 100, or 50 BBUs/Cells). The average number of users per cell was 30, moving at various speed ranges: 8-16, 16-25, 25-33, 33-50 m/s. In addition, the minimum rate requirement for cellular users was assumed to be 36 kbps [70]. As per [70], it was shown that a typical VoLTE call generates a traffic with an equivalent minimum rate requirement of 36 kbps. Therefore, this work assumes that each user has a minimum rate requirement of 36 kbps in each scheduling time slot. Since this work tackles the problem at the bit level, it is assumed that the rate requirement must be satisfied in each time slot for all users. Moreover, the number of physical machines used was ten, each of which represents a pool of BBUs. The BBUs were clustered as virtual machines in each physical machine. A different number of virtual machines was used for each configuration; more specifically 15, 10, or 5, respectively. Table 5.2 summarizes the simulation parameters of the system.

### 5.4.2 Results:

#### 5.4.2.1 Computational Resource Consumption

Figure 5.3 shows the aggregated computational resource consumption of the different BBUs for the three considered configurations at different speeds. Several observations can be made. Our first observation is that the average computational resource consumption decreases as the speed increases. This is to be expected since the channel condition becomes worse as the users' speed increases. This results in a worse SNR and hence lower computational requirements. Our second observation is that as the number of BBUs increases, the better distribution of the required computational resources among the BBUs results in a more balanced consumption. Consequently, the packing efficiency

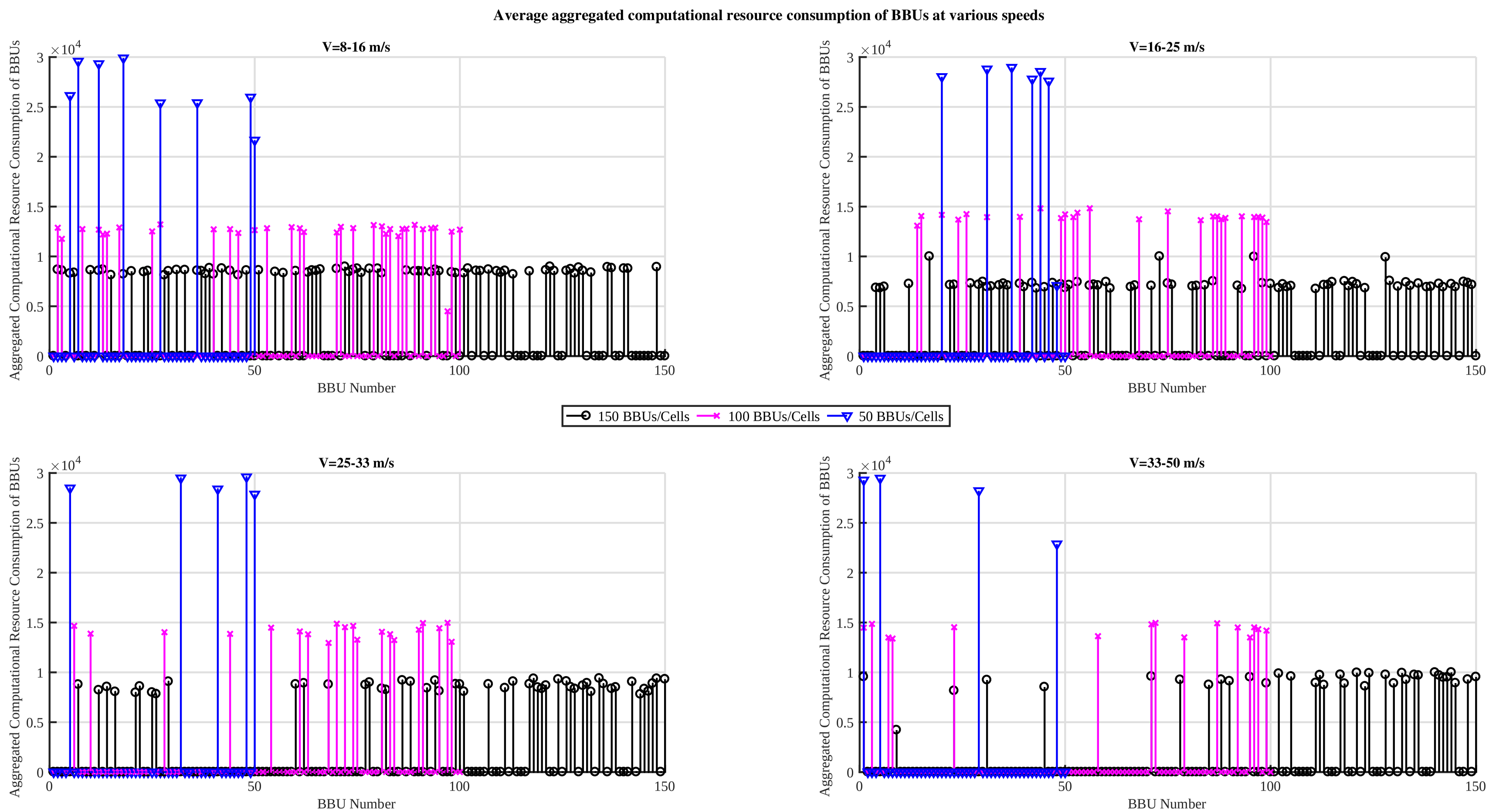


Figure 5.3: Aggregated computational resource consumption of BBUs at various speeds.

Table 5.2: Simulation Parameters & Values

| Parameter | Value |
|---|---|
| Carrier frequency | 2 GHz |
| Bandwidth | 10 MHz |
| Number of RBs | 50 |
| Number of subcarriers per RB | 12 |
| Subcarrier spacing | 15 KHz |
| RB Bandwidth | 180 KHz |
| eNodeB Tx power | 20 W |
| Slot duration | 1 ms |
| Average Number of Cellular Users | 30 |
| Cell-level user distribution | Uniform |
| Log-normal shadowing standard deviation | 8 dB |
| $\theta$ | 8 |
| $L_B$ | 58 |
| $w$ | 3.75 W |
| $P_{st}^{j}$ | 1.25 W |
| $P_{st}^{pool}$ | 200 W |

improves, since it becomes easier to pack the items into the BBUs. This is so because having a smaller BBU size makes it easier to allocate the computational resources among the different BBUs.

Figure 5.4 shows the average BBU computational resource consumption at different speeds for both the decomposition model and the heuristic model. The first observation that can be made is that the computational resource consumption decreases as the speed increases. This follows from the previous figure in which the same observation was made and is illustrated by the fact that a lower number of BBUs was needed to satisfy the requirements as the speed increased. This is also evident for the data rate as it decreases when the speed increases as noticed in the right hand figure. This is because higher speeds is synonymous with worse channel conditions and hence lower data rate. It can also be seen that both the BIP-based solution and heuristic solution achieve a much higher average data rate than the considered threshold for the different configurations at various speeds. This illustrates that the data rate is indeed maximized as per the formulated optimization problem. The second observation is that the heuristic method achieves close to optimal performance. This coincides with the results from [67] in which the greedy algorithm was shown to have a close to optimal performance when compared to the BIP-based solution. Hence, the computational requirements of the heuristic so-

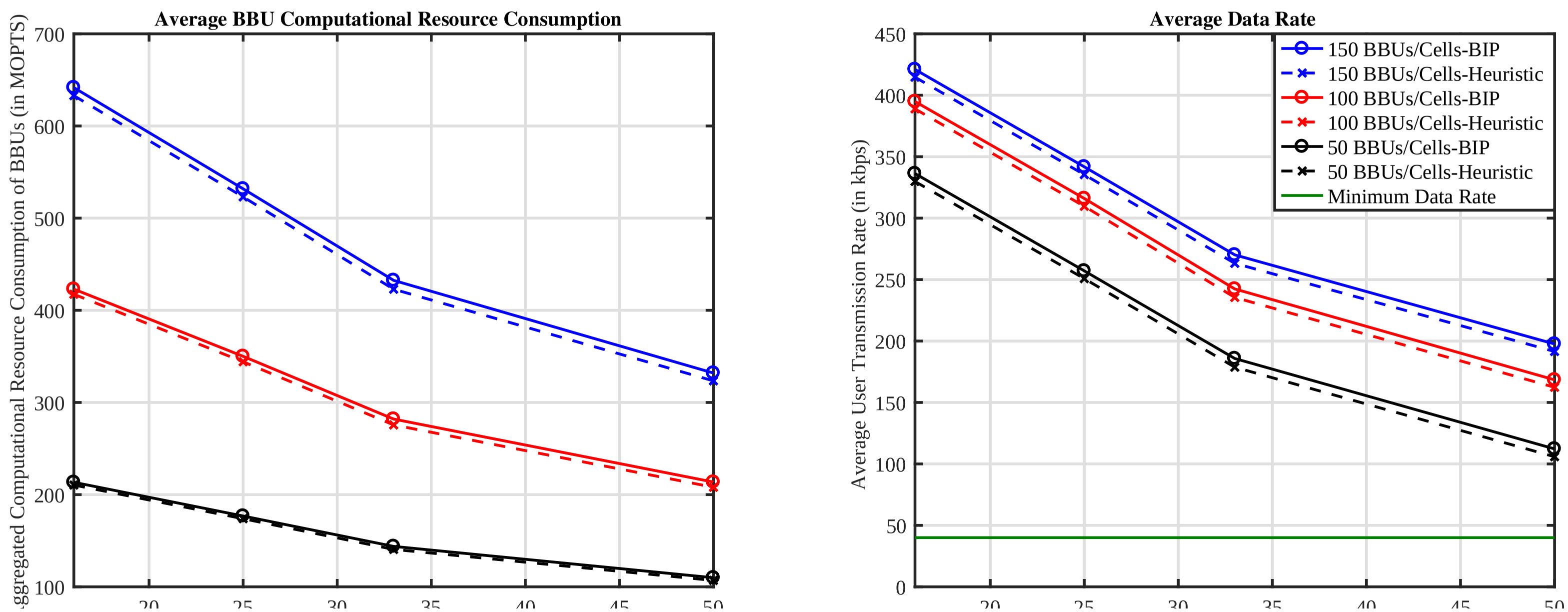


Figure 5.4: BBU computational resource consumption at various speeds.

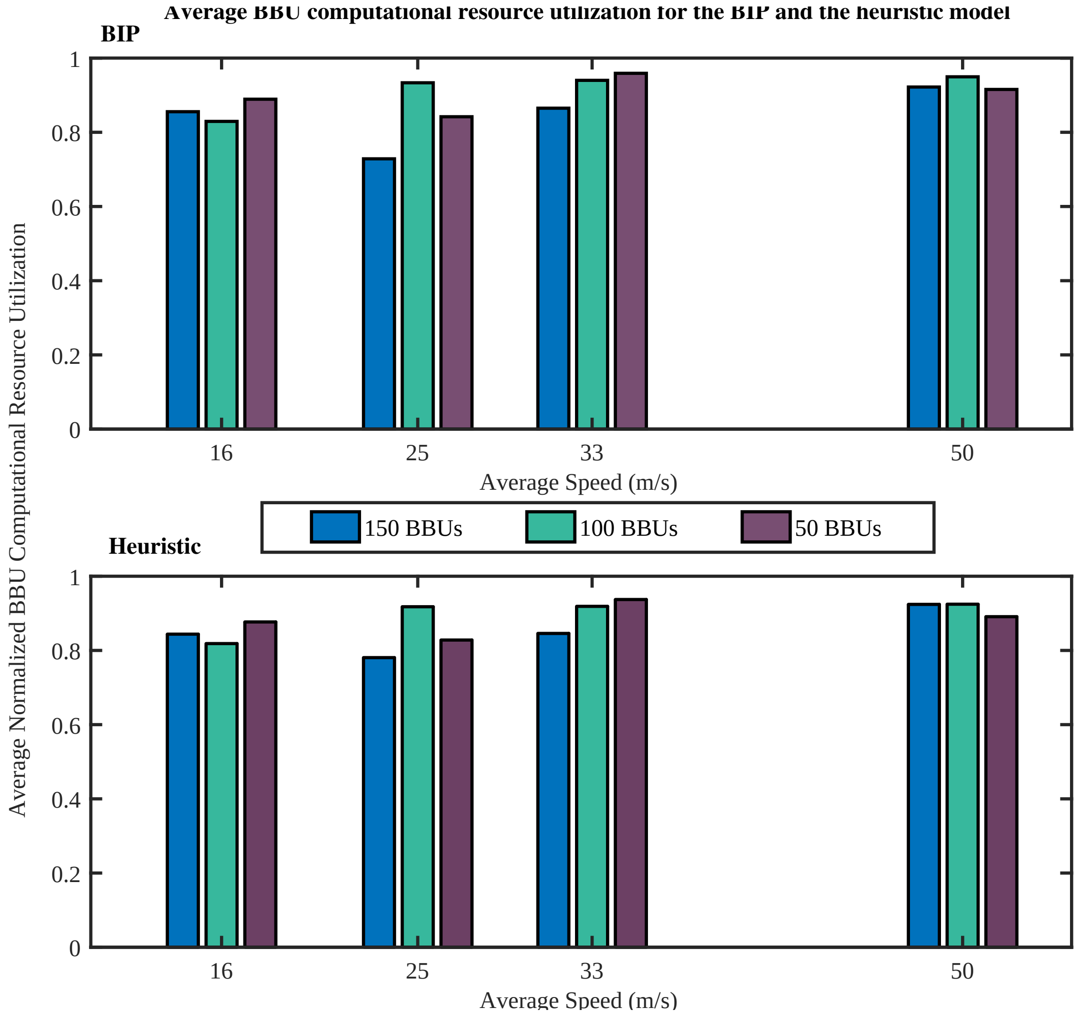


Figure 5.5: Average BBU computational resource utilization for the BIP and the heuristic models.

lution are expected to be close to those of the BIP-based solution since the resource allocation of both solutions yielded similar results.

Figure 5.5 shows the average BBU computational resource utilization. It should be noted that both the BIP and the heuristic models perform efficiently at different speeds. The average utilization of the BBU is between approximately 80 to 95%, which indicates a highly efficient allocation in both the BIP and the heuristic models. This results in a lower power consumption for both the BBUs and the physical machines.

#### 5.4.2.2 Power Consumption

Power consumption of the system is an important metric to be investigated. It is worth noting that the power consumption is considered in this work rather than the energy consumption because service providers prefer to have a holistic overview of their system's performance. Hence, the average consumption over time is of more interest to the service provider than the instantaneous consumption. Therefore, the power consumption of the different configurations is evaluated. In particular, the average BBU power consumption and that of the physical machines used to host the BBUs is shown below.

Figure 5.6 shows the average BBU power consumption at different speeds. Similar to the trend in Figure 5.4, the average power consumption of the BBUs decreases as speed increases. This is because the computational resources decreased and therefore less power is needed to perform the required computations. Furthermore, it can be seen that the heuristic solution achieved close to optimal performance here as well. This was expected since it is known that the FFD bin packing algorithm achieves close to optimal results. It is worth noting that the heuristic solution consumes less power since the computational resource consumption is lower. This does not mean that the heuristic solution outperforms the BIP-based solution in terms of power, but rather that the BIP-based solution yields a better resource allocation that results in higher computational resource requirements which is translates to a higher power consumption.

Figure 5.7 shows the average power consumption of the physical machines used to host the BBUs. A similar trend to that observed in Figs. 5.4 and 5.6 can be seen in this case. Since the computational resource consumption of the BBUs decreases as the speed increases, less physical machines are required to host these BBUs. Thus, less power is needed by the physical machines to host the BBUs serving the cells within the system. Moreover, the heuristic solution was seen to achieve close to optimal results similar to those seen in Figure 5.6.

#### 5.4.2.3 Scalability Analysis

The scalability of the system has been analyzed in terms of aggregated computational resource consumption and corresponding power consumption. Two parameters have been considered, namely the number of RBs and the average number of users.

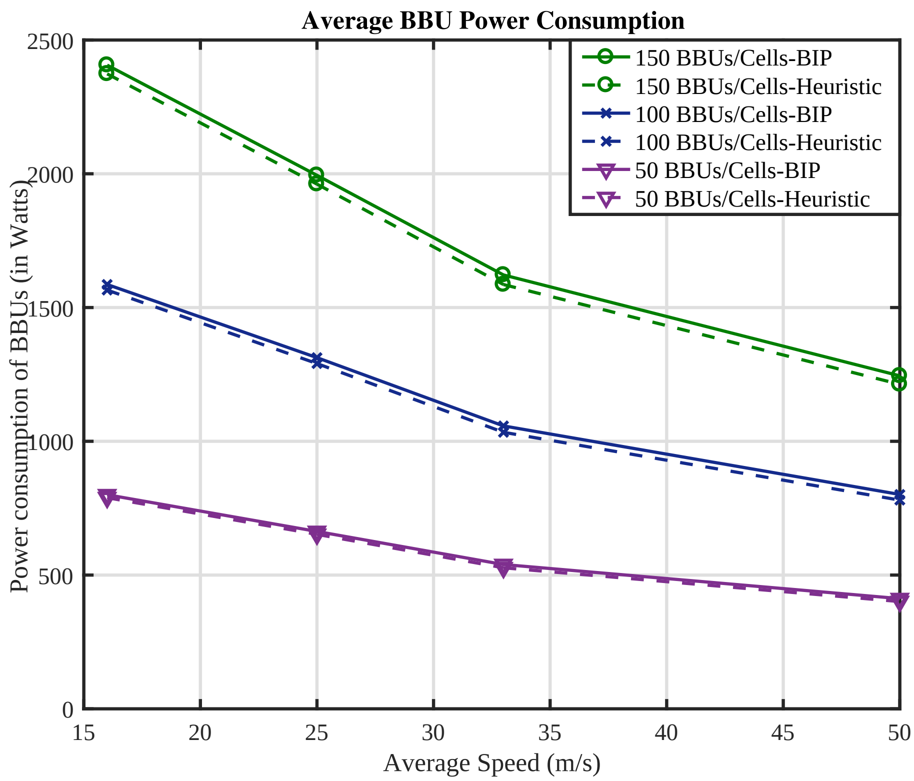


Figure 5.6: BBU power consumption at various speeds.

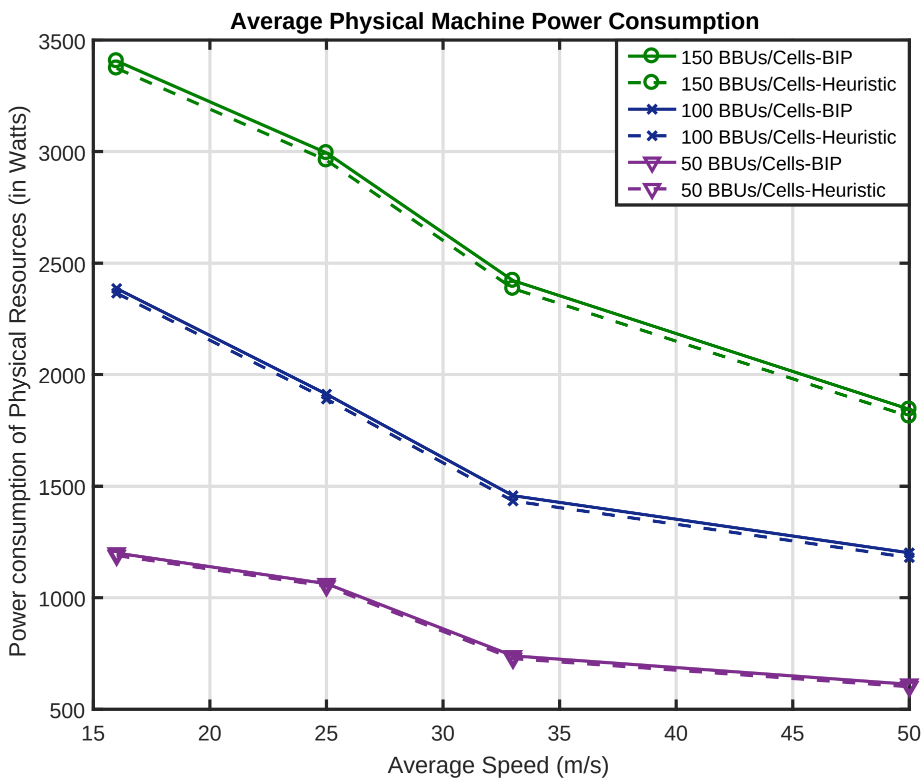


Figure 5.7: Physical machines' power consumption at various speeds.

Figure 5.8 shows the average computational resource and power consumption of the 150 BBUs for different number of RBs. Similar to previous figures, it is noticed that as the speed increases, the computational resource and power consumption decreases. Moreover, it is noticed that as the number of RBs increases, the average computational resource consumption increases and consequently the average power consumption increases. This is expected since more RBs means more resources can be allocated to users. This leads to users transmitting at higher rates which results in the higher computational resource and power consumption. Furthermore, it is observed that the increase is linear, which supports the scalability of the system. Figure 5.9 shows the average computational resource and power consumption of the 150 BBUs for different average number of users. The same trend in terms of the computational resource and power consumption is observed. Additionally, it can be noticed that as the speed increases, the rate of increase of both computational resource and power consumption becomes higher. On top of that, the figure shows that as the average number of users increases, the considered parameters increase in a logarithmic manner. This further cements the notion that the system is scalable.

## 5.5 Conclusion

C-RAN technology is a new paradigm that changes the way wireless networks work with the help of cloud computing that enables wireless operators to overcome two of their most major concerns, namely scalability and availability. In addition, cloud computing enables implementing other futuristic aspects such as serving heterogeneous networks under one umbrella. Despite the countless advantages that cloud computing provides, it is probable that a large waste of resources may occur due to the lack of optimal allocation procedures, consequently resulting in significant power losses in the system. This would be particularly serious when the resources are intended to serve large scale systems containing dense, small cells. As a result, optimizing resource allocation is crucial to achieve high efficiency. In this work, an allocation problem was modeled to determine the optimal allocation of resources between RRH and BBU. In addition, the computational resources required by each user were determined by modeling the assignment of the physical resources of the RRHs to the users. The formulated problem is one of

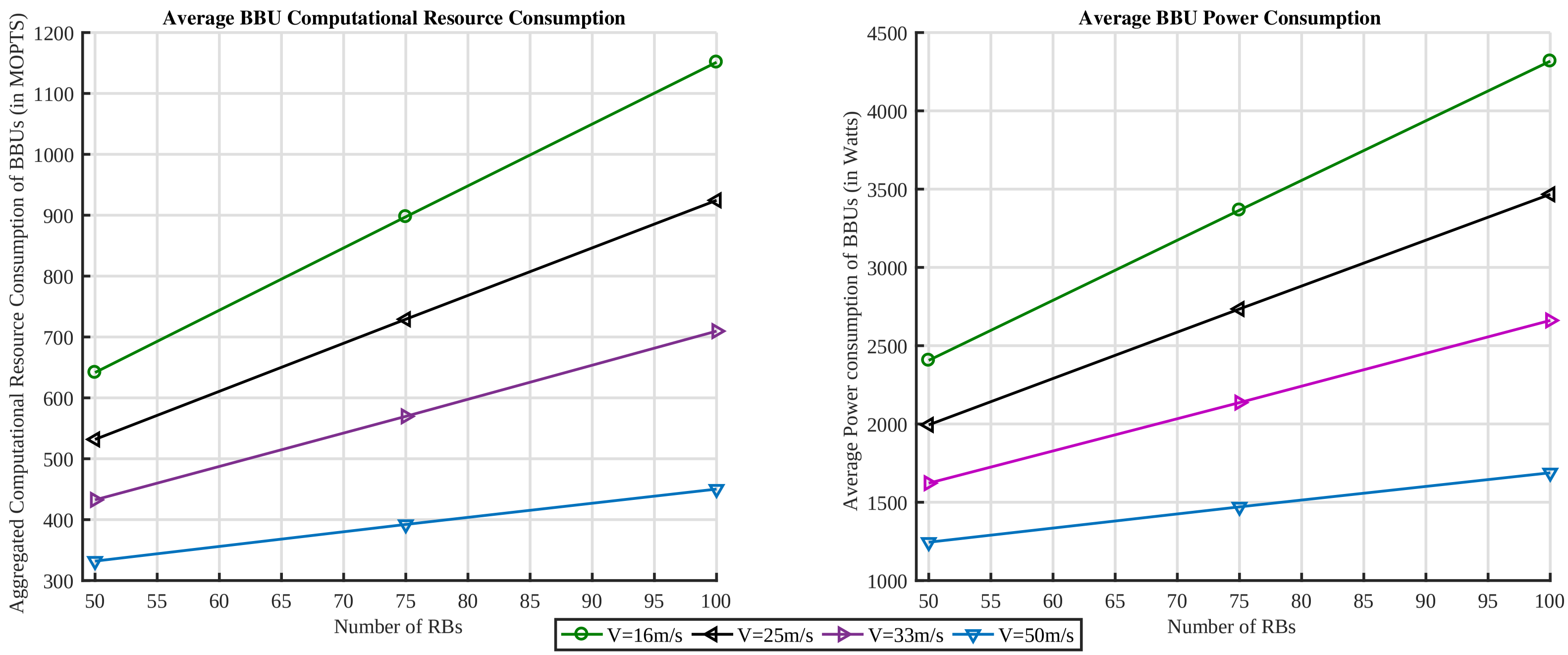


Figure 5.8: Average BBU computational resource and BBU power consumption for different number of RBs.

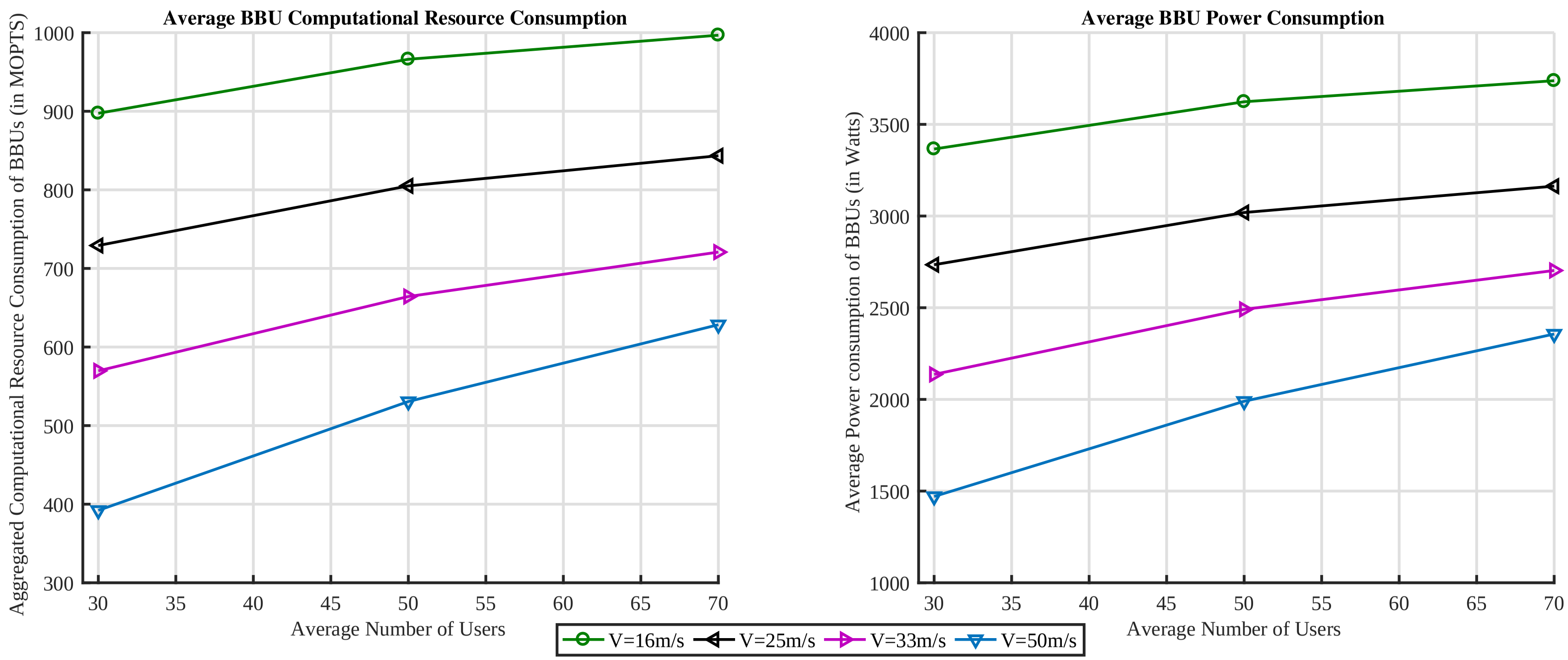


Figure 5.9: Average BBU computational resource and BBU power consumption for different number of users.

high complex problems and was solved using the proposed decomposition model. It was noted that the proposed model is highly efficient for BBU allocation. Furthermore, the power consumption and the computational resource consumption of the BBUs decreases as the channel worsens. Lastly, a heuristic solution was developed to follow the same criteria of decomposition model and has been observed that it can offer close to optimal performance with less power consumption.

# Chapter 6
# Conclusion

## 6.1 Introduction

This chapter summarizes the contributions of this thesis and the conclusions drawn from the results obtained.

## 6.2 Summary of Contributions

Wireless networks are facing various challenges that demand continuous and rapid improvement. Long-Term Evolution (LTE) is a preferred wireless technology because of its satisfactory performance. Owing to an exponential increase in demand and new potential applications, the core network of LTE, which is known as the Evolved Packet Core (EPC), is affected by a surge in signaling caused by a variety of control functions. The signaling overhead decreases the users' Quality of Experience (QoE). Furthermore, the rapid growth of data traffic triggers the demand of innovative paradigm such as C-RAN technology. C-RAN has influenced researchers in both academia and industry to bring forth a new era of cellular networks. However, the dense C-RAN architecture imposes several concerns in terms of power consumption whereby optimization techniques are needed to enhance the overall performance. This thesis aims to present practical models for two sides namely; the signaling overhead caused by the location management techniques and optimizing the power consumption resulted in the rapid growth of the data traffic in the C-RAN environment.

Chapter two tackled the problem of the signaling overhead for both uplink and downlink. An efficient and novel algorithm that enables a significant reduction of the TAU and paging signaling traffic was defined. To the best of our knowledge, the proposed design is the first that includes MME in the system optimization. Two schemes were investigated, namely, the distributed MME scheme and the centralized MME scheme. In these schemes, both overlapping TAL assignment and efficient mapping of the MME to the TA/TAL are considered. Furthermore, a heuristic algorithm is presented as a low-complexity approach that gives a sub-optimal solution.

Chapter three intended to enable adaptive online cell-to-TAL assignment in order to further investigate the proposed pooling schemes.Thus, cell-to-TAL assignment can be engineered dynamically while the UE is in continuous movement. The system keeps analyzing the mobility pattern and continuously updates the TA assigned to the list. Thus, the frequency of TAU will be reduced significantly. The proposed dynamic technique was realized through an SON scheme along with a new smart cell selection approach instead of the conventional ring-based cell selection presented in the literature. A new heuristic algorithm differs from the one proposed in the previous chapter and constitutes of two sub-problems in the same manner as the decomposition model was proposed. The algorithm dynamically diversifies the TALs among the cells which helps in reducing the TAU signaling load. The results showed that the centralized scheme outperforms the distributed one. Also, the heuristic algorithm offered sub-optimal results when compared to the LP solution.

Chapter three introduced three evolutionary artificial algorithms namely; Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), and Gravitational Search Algorithm (GSA), in order to solve the problem of minimizing the signaling overhead. It was shown that ABC-based deployment is faster in convergence and can improve the minimization of signaling better than other optimization heuristics considered. Moreover, the measured relative standard deviation (RSD) value of the applied algorithms shows low uncertainty of around 1% for the objective function and 3% for the paging, tracking area update, and power. Hence, the three applied optimization algorithms have proven to be efficient and precise for solving the problem in a large scale environment.

Chapter four formulated an optimization model depicting the two levels of scheduling (from physical resources to users and from RRHs to BBUs) in a C-RAN environment. In the level comprised between cells and users, resources are distributed among users, which have different QoS requirements. As a consequence, the system has to optimize resource allocation accordingly while maintaining other aspects such as availability of physical resources, satisfying the QoS, and continuity of service. In the RRH and BBU level, computing requirements need to be processed instantly in the available BBU pool available while maintaining power consumption and optimizing computing resources. It was observed that the computational resource requirements as well as the power consumption of the BBUs and the physical machines decrease as the channel quality worsens. Moreover, the developed heuristic solution achieved a close to optimal performance while having a lower complexity. Lastly, the average utilization of the BBU's resources for both the BIP and the heuristic is between 80 to 95%, which indicates a highly efficient utilization of the proposed solutions.

## 6.3 Future Work

The work of this thesis can be extended in several directions. For instance, a methodology can be introduced to adapt the proposed heuristic algorithms dynamically in order to mimic realistic networks with various number of cells and different mobility patterns with the same aim being to minimize the signaling overhead. Another approach is to incorporate a framework that has the ability to compress the signaling traffic as well as to suitably allocate the cells to MMEs.

Another direction that can be investigated is to consider a heterogeneous user traffic model and study its impact on both the computational requirement and power consumption in a C-RAN environment. This allows for a better understanding and portrayal of real life networks under different quality of service requirements.

# References

[1] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the sixth international symposium on micro machine and human science*, vol. 1. New York, NY, 1995, pp. 39–43.

[2] M. A. Abido, "Optimal power flow using particle swarm optimization," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 7, pp. 563–571, 2002.

[3] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm," *Journal of global optimization*, vol. 39, no. 1, pp. 459–471, 2007.

[4] H. A. Hashim, B. Ayinde, and M. Abido, "Optimal placement of relay nodes in wireless sensor network using artificial bee colony algorithm," *Journal of Network and Computer Applications*, vol. 64, pp. 239–248, 2016.

[5] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Information sciences*, vol. 179, no. 13, pp. 2232–2248, 2009.

[6] "Signaling is growing 50% faster than data traffic," Nokia Siemens Network, 2012.

[7] T. Deng, X. Wang, P. Fan, and K. Li, "Modeling and performance analysis of tracking area list-based location management scheme in lte networks," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[8] "New son use-case: Tracking area optimization,," 3GPP TSG-RAN WG3 Std. R3-071 594, Aug 2007.

[9] J. Ferragut and J. Mangues-Bafalluy, "A self-organized tracking area list mechanism for large-scale networks of femtocells," in *2012 IEEE International Conference on Communications (ICC'12)*, June 2012, pp. 5129–5134.

[10] H. A. Hashim, S. El-Ferik, and M. A. Abido, "A fuzzy logic feedback filter design tuned with pso for l1 adaptive controller," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9077–9085, 2015.

[11] A. Hatamlou, S. Abdullah, and H. Nezamabadi-Pour, "A combined approach for clustering based on k-means and gravitational search algorithms," *Swarm and Evolutionary Computation*, vol. 6, pp. 47–52, 2012.

[12] R. E. Brown, E. R. Masanet, B. Nordman, W. F. Tschudi, A. Shehabi, J. Stanley, J. G. Koomey, D. A. Sartor, and P. T. Chan, "Report to congress on server and data center energy efficiency: Public law 109-431," 06/2008 2008.

[13] M. Dabbagh, B. Hamdaoui, M. Guizani, and A. Rayes, "Toward energy-efficient cloud computing: Prediction, consolidation, and overcommitment," *IEEE Network*, vol. 29, no. 2, pp. 56–61, March 2015.

[14] ——, "Energy-efficient cloud resource management," in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2014, pp. 386–391.

[15] "The impact of small cells on mme signaling," Alcatel-Lucent, 2013.

[16] "Managing the signaling traffic in packet core," Alcatel-Lucent, 2012.

[17] K. W. Tindell, A. Burns, and A. J. Wellings, "Allocating hard real-time tasks: An np-hard problem made easy," *Real-Time Systems*, vol. 4, no. 2, pp. 145–165, 1992. [Online]. Available: `http://dx.doi.org/10.1007/BF00365407`

[18] M. R. Garey and D. S. Johnson, *Computers and intractability.* wh freeman New York, 2002, vol. 29.

[19] M. Kalil, A. Shami, and Y. Ye, "Wireless resources virtualization in LTE systems," in *IEEE Conference on Computer Communications Workshops (INFOCOM '14 WKSHPS)*, April 2014, pp. 363–368.

[20] I. Widjaja, P. Bosch, and H. La Roche, "Comparison of MME signaling loads for long-term-evolution architectures," in *2009 IEEE 70th Vehicular Technology Conference Fall (VTC'09-Fall)*, Sept 2009, pp. 1–5.

[21] H. Hawilo, A. Shami, M. Mirahmadi, and R. Asal, "NFV: state of the art, challenges, and implementation in next generation mobile networks (vepc)," *IEEE Network*, vol. 28, no. 6, pp. 18–26, Nov 2014.

[22] S. Razavi, D. Yuan, F. Gunnarsson, and J. Moe, "Exploiting tracking area list for improving signaling overhead in LTE," in *2010 IEEE 71st Vehicular Technology Conference (VTC'10-Spring)*, May 2010, pp. 1–5.

[23] S.-R. Yang, Y.-C. Lin, and Y.-B. Lin, "Performance of mobile telecommunications network with overlapping location area configuration," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1285–1292, March 2008.

[24] D. Gu and S. Rappaport Stephen, "Mobile user registration in cellular systems with overlapping location areas," in *1999 IEEE 49th Vehicular Technology Conference (VTC'99)*, vol. 1, Jul 1999, pp. 802–806 vol.1.

[25] K.-H. Chiang and N. Shenoy, "A 2-d random-walk mobility model for location-management studies in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 2, pp. 413–424, March 2004.

[26] ——, "Performance of an overlapped macro-cell based location area scheme," in *IEEE International Conference on Communications (ICC '03)*, vol. 1, May 2003, pp. 475–481 vol.1.

[27] R.-H. Liou, Y.-B. Lin, and S.-C. Tsai, "An investigation on LTE mobility management," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 166–176, Jan 2013.

[28] S. Razavi, D. Yuan, F. Gunnarsson, and J. Moe, "Dynamic tracking area list configuration and performance evaluation in LTE," in *2010 IEEE GLOBECOM Workshops (GC'10 Wkshps)*, Dec 2010, pp. 49–53.

[29] S. Modarres Razavi and D. Yuan, "Mitigating mobility signaling congestion in LTE by overlapping tracking area lists," in *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '11. New York, NY, USA: ACM, 2011, pp. 285–292. [Online]. Available: `http://doi.acm.org/10.1145/2068897.2068947`

[30] S. Razavi and D. Yuan, “Reducing signaling overhead by overlapping tracking area list in LTE,” in *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC’14)*, May 2014, pp. 1–7.

[31] S. Ikeda, N. Kami, and T. Yoshikawa, “A tracking area list configuration method to mitigate burst location updates,” in *2014 IEEE Fifth International Conference on Communications and Electronics (ICCE’14)*, July 2014, pp. 258–263.

[32] Y. W. Chung, “Adaptive design of tracking area list in LTE,” in *2011 Eighth International Conference on Wireless and Optical Communications Networks (WOCN’11)*, May 2011, pp. 1–5.

[33] T. Taleb, K. Samdanis, and A. Ksentini, “Supporting highly mobile users in cost-effective decentralized mobile operator networks,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 7, pp. 3381–3396, Sept 2014.

[34] A. Kunz, T. Taleb, and S. Schmid, “On minimizing serving GW/MME relocations in LTE,” in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, ser. IWCMC ’10. New York, NY, USA: ACM, 2010, pp. 960–965. [Online]. Available: http://doi.acm.org/10.1145/1815396.1815615

[35] K. Kyamakya and K. Jobmann, “Location management in cellular networks: classification of the most important paradigms, realistic simulation framework, and relative performance analysis,” *IEEE Transactions on Vehicular Technology*, vol. 54, no. 2, pp. 687–708, March 2005.

[36] E. Aqeeli, A. Moubayed, and A. Shami, “Towards intelligent LTE mobility management through MME pooling,” in *IEEE 58th Global Communications Conference (GLOBECOM’15)*, Dec. 2015.

[37] S. Modarres Razavi, D. Yuan, F. Gunnarsson, and J. Moe, “On dynamic signaling congestion mitigation by overlapping tracking area lists,” *J. Netw. Comput. Appl.*, vol. 53, no. C, pp. 164–172, Jul. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.jnca.2015.03.009

[38] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile users: to update or not to update?" in *13th IEEE Networking for Global Communications (INFOCOM '94)*, Jun 1994, pp. 570–576 vol.2.

[39] G. Wan and E. Lin, "A dynamic paging scheme for wireless communication systems," in *3rd Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ser. MobiCom '97. New York, NY, USA: ACM, 1997, pp. 195–203. [Online]. Available: http://doi.acm.org/10.1145/262116.262147

[40] S. Tabbane, "An alternative strategy for location tracking," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 5, pp. 880–892, Jun 1995.

[41] M. Toril, S. Luna-Ramirez, and V. Wille, "Automatic replanning of tracking areas in cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2005–2013, Jun 2013.

[42] P. Bonami, L. T. Biegler, A. R. Conn, G. CornuéJols, I. E. Grossmann, C. D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, and A. WäChter, "An algorithmic framework for convex mixed integer nonlinear programs," *Discret. Optim.*, vol. 5, no. 2, pp. 186–204, May 2008. [Online]. Available: http://dx.doi.org/10.1016/j.disopt.2006.10.011

[43] "Self-optimizing network, the benefits of son in lte,," http://www.4gamericas.org/ 2011.

[44] P. R. L. Gondim, "Genetic algorithms and the location area partitioning problem in cellular networks," in *Proceedings of Vehicular Technology Conference - VTC*, vol. 3, Apr 1996, pp. 1835–1838 vol.3.

[45] J. Taheri and A. Y. Zomaya, "A genetic algorithm for finding optimal location area configurations for mobility management," in *The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05)l*, Nov 2005, pp. 9 pp.–577.

[46] ——, "A combined genetic-neural algorithm for mobility management," in *Proceedings 20th IEEE International Parallel Distributed Processing Symposium*, April 2006, pp. 8 pp.–.

[47] L. Wang and G. Si, "Optimal location management in mobile computing with hybrid genetic algorithm and particle swarm optimization (ga-pso)," in *2010 17th IEEE International Conference on Electronics, Circuits and Systems*, Dec 2010, pp. 1160–1163.

[48] N. Ahmad, H. Safa, and W. El-Hajj, "On the tas reconfiguration problem in lte networks," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Sept 2016, pp. 463–468.

[49] J. Du, L. Zhao, J. Xin, J.-M. Wu, and J. Zeng, "Using joint particle swarm optimization and genetic algorithm for resource allocation in td-lte systems," in *2015 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE)*, Aug 2015, pp. 171–176.

[50] J. Du, L. Zhao, J. Feng, J. Xin, and Y. Wang, "Enhanced pso based energy-efficient resource allocation and cqi based mcs selection in lte-a heterogeneous system," *China Communications*, vol. 13, no. 11, pp. 197–204, Nov 2016.

[51] L. Huang, X. Chen, Z. Gao, and H. Cai, "Online propagation model correction based on pso algorithmin lte son system," in *Anti-counterfeiting, Security, and Identification*, Aug 2012, pp. 1–4.

[52] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Erciyes university, computer engineering department, engineering faculty, Tech. Rep., 2005.

[53] H. A. Hashim and M. Abido, "Fuzzy controller design using evolutionary techniques for twin rotor mimo system: a comparative study," *Computational intelligence and neuroscience*, vol. 2015, no. 9, p. 11, 2015.

[54] Y. Liao, L. Song, Y. LI, and Y. Zhang, "How much computing capability is enough to run a cloud radio access network?" *IEEE Communications Letters*, vol. PP, no. 99, pp. 1–1, 2016.

[55] Y. Du and G. de Veciana, ""'wireless networks without edges": Dynamic radio resource clustering and user scheduling," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 1321–1329.

[56] X. Chen, N. Li, J. Wang, C. Xing, L. Sun, and M. Lei, "A dynamic clustering algorithm design for c-ran based on multi-objective optimization theory," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, May 2014, pp. 1–5.

[57] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-ran," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[58] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Quality of service aware dynamic bbu-rrh mapping in cloud radio access network," in *2015 International Conference on Emerging Technologies (ICET)*, Dec 2015, pp. 1–5.

[59] D. Mishra, P. C. Amogh, A. Ramamurthy, A. A. Franklin, and B. R. Tamma, "Load-aware dynamic rrh assignment in cloud radio access networks," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[60] D. Zhu and M.Lei, "Traffic adaptation and energy saving potential of centralized radio access networks with coordinated resource allocation and consolidation," in *2013 8th International Conference on Communications and Networking in China (CHINACOM)*, Aug 2013, pp. 587–593.

[61] D.Zhu and M.Lei, "Traffic and interference-aware dynamic bbu-rru mapping in c-ran tdd with cross-subframe coordinated scheduling/beamforming," in *2013 IEEE International Conference on Communications Workshops (ICC)*, June 2013, pp. 884–889.

[62] O. Dhifallah, H. Dahrouj, T. Y. Al-Naffouri, and M. S. Alouini, "Joint hybrid backhaul and access links design in cloud-radio access networks," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Sept 2015, pp. 1–5.

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

[64] Q. Liu, G. Wu, Y. Guo, Y. Zhang, and S. Hu, "Energy efficient resource allocation for control data separated heterogeneous-cran," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[65] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo, "Cloudiq: A framework for processing base stations in a data center," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 125–136. [Online]. Available: http://doi.acm.org/10.1145/2348543.2348561

[66] M. Qian, W. Hardjawana, J. Shi, and B. Vucetic, "Baseband processing units virtualization for cloud radio access networks," *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 189–192, April 2015.

[67] A. Moubayed, A. Shami, and H. Lutfiyya, "Wireless Resource Virtualization With Device-to-Device Communication Underlaying LTE Network," *IEEE Transactions on Broadcasting*, vol. 61, no. 4, pp. 734–740, Dec 2015.

[68] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in c-ran," in *2015 European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 169–174.

[69] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, October 2011.

[70] E. Jailani, M. Ibrahim, and R. Rahman, "LTE speech traffic estimation for network dimensioning," in *IEEE Symposium on Wireless Technology and Applications (ISWTA'12)*, Sept 2012, pp. 315–320.

# Curriculum Vitae

## VITA

| | |
|---|---|
| **Name:** | **Emad Ali Aqeeli** |
| **Post-secondary: Education and** | **King Abdulaziz University**<br>**B.Sc. in Electrical-Computer Engineering**<br>Jeddah, Kingdom of Saudi Arabia<br>2001-2006 |
| | **King Abdulaziz University**<br>**M.Sc. in Electrical-Computer Engineering**<br>Jeddah, Kingdom of Saudi Arabia<br>2008-2011 |
| | **The University of Western Ontario**<br>**Ph.D. in Electrical-Computer Engineering**<br>London, Ontario, Canada<br>2014-2017 |
| **Related Work Experience** | **Teaching Assistant**<br>The University of Western Ontario<br>2014-2017 |

## Publications:

- **Towards Intelligent LTE Mobility Management through MME Pooling**

  IEEE Global Communications Conference (GLOBECOM), 2015

- **Dynamic SON-Enabled Location Management in LTE Network**

  IEEE Transactions on Mobile Computing, 2017

- **Power-Aware Optimized RRH to BBU Allocation in C-RAN**

  IEEE Transactions on Wireless Communications, 2017