RRH-Sector selection and load balancing based on MDP and Dynamic RRH-Sector-BBU Mapping in C-RAN

by

Mostafa MOUAWAD

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL,13 DECEMBER, 2021

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Zbigniew Dziong, Thesis Supervisor
Department of Electrical Engineering, École de Technologie Supérieure

Mr. Christopher Fuhrman, President of the Board of Examiners
Department of Software and IT Engineering, École de Technologie supérieure

Mr. Michel Kadoch, Member of the jury
Department of Electrical Engineering, École de Technologie Supérieure

Mr. Wessam Ajib, External Independent Evaluator
Department D'Informatique, Universite du Quebec a Montreal

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC

MONTREAL, 24 NOVEMBER, 2021

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGMENT

Firstly, It was my pleasure to be supervised by Professor Zbigniew Dziong, I want to thank him for his unlimited assistance, valuable guidance, and recommendations through my Ph.D. research.

I would like to express my most profound gratitude to my parents, my sister, and my lovely friends for their unconditional love, support and great patience all over my whole life.

Also, I would like to thank Brunel university in London, and especially Wireless Networks and Communications Centre (WNCC), for allowing me to do an exchange during Ph.D. studies.

Last, but by no means least, I would like to thank my close friends Ahmed Emam, and Khaled Addali for their support and help during my Ph.D. journey.

Sélection du secteur RRH et équilibrage de charge sur la base du MDP et du mappage dynamique RRH-Secteur-BBU dans C-RAN

Mostafa MOUAWAD

RÉSUMÉ

L'augmentation constante du trafic dans les réseaux sans fil oblige les opérateurs à améliorer considérablement leur infrastructure de réseau. Cependant, les dépenses prévues pour la construction, l'exploitation et la mise à jour du réseau d'accès radio (RAN) imposent un défi puisque le revenu moyen par utilisateur (ARPU) est presque constant ou diminue lentement. De plus, la nécessité de connecter des appareils sans fil a provoqué une augmentation du volume du trafic de données mobiles. Par conséquent, les opérateurs mobiles doivent chercher des solutions qui peuvent fournir un équilibre entre atteindre un niveau de service satisfaisant pour les clients et les coûts, pour optimiser les bénéfices et la croissance. Le réseau d'accès radio en nuage (C-RAN) est une nouvelle architecture qui prend en charge l'augmentation considérable du trafic du réseau mobile. Malgré des gains offertes par C-RAN, un environnement de trafic variable dans le temps peut provoquer des déséquilibres de charge, entraînant une utilisation inefficace des ressources. Par conséquent, les performances du réseau peuvent se dégrader en termes d'utilisateurs bloqués, de nombre de transferts inutiles et de consommation d'énergie. Cette thèse présente une sélection de paires RRH-secteur pour de nouvelles connexions et un cadre d'équilibrage de charge réseau qui optimise la qualité de service (QoS) et la récompense de l'opérateur dans le C-RAN. Dans la première partie de la thèse, nous proposons un nouvel algorithme de sélection RRH-Secteur où il sélectionne la meilleure paire RRH-Secteur pour chaque nouvelle demande de connexion en tenant compte des objectifs de l'utilisateur et de l'opérateur. Cette décision est basée sur un algorithme dérivé du processus décisionnel de Markov (MDP) dans le but de maximiser l'utilité intégrée opérateur-utilisateur. Dans la deuxième partie du cadre, le problème d'équilibrage de charge est abordé via l'optimisation de le mappage dynamique RRH-Secteur-BBU formulée comme un problème d'optimisation sous contrainte linéaire basé sur des nombres entiers. Nous comparons les solutions à ce problème obtenues par plusieurs algorithmes évolutionnaires tels que : colonie d'abeilles (BCO), recherche de Cuckoo (CUCO), algorithmes génétiques (GA) et essaim de particules (PSO). Enfin, nous évaluons les solutions proposées à l'aide de simulations approfondies. Les résultats montrent que le schéma de sélection du secteur RRH proposé offre des gains significatifs en termes de récompense de l'opérateur et de probabilité de blocage de la connexion par rapport à la méthode d'intensité du signal reçu (RSS). De plus, les algorithmes évolutifs sont comparés à la méthode de recherche exhaustive qui donne le mappage optimale RRH-Secteur-BBU. Les résultats montrent que, dans la plupart des scénarios considérés, les algorithmes proposés atteignent les solutions optimales en termes de nombre d'utilisateurs bloqués, de nombre de transferts d'utilisateurs entre les secteurs, et de consommation d'énergie BBU. Par conséquent, la qualité de service est maximisée et le réseau équilibré est obtenu.

Mots-clés: Réseau radio d'accès infonuagique, équilibrage de charge, processus de décision de Markov, QoS

RRH-Sector selection and load balancing based on MDP and Dynamic RRH-Sector-BBU Mapping in C-RAN

Mostafa MOUAWAD

ABSTRACT

The ever-rising level of traffic in wireless networks is forcing operators to improve significantly their network infrastructure. However, the expected expenditure for building, operating and updating the Radio Access Network (RAN) imposes a challenge since the Average Revenue per User (ARPU) is almost constant or declining slowly. Consequently, mobile operators need to search for solutions that can provide a balance between reaching a satisfactory level of service to customers and costs to optimize the profit and growth. Cloud radio access network (C-RAN) is a novel architecture that supports the tremendous increase in mobile network traffic. Despite the gains that C-RAN offers, a time-varying traffic environment can cause load imbalances, resulting in inefficient resource utilization. Consequently, the network performance can degrade in terms of the blocked users, the number of unnecessary handovers, and the power consumption. This thesis presents an RRH-Sector pair selection for new connections and network load-balancing framework that optimizes the Quality of Service (QoS) and operator reward in C-RAN. In the first part of the framework, we propose a novel RRH-Sector selection algorithm that selects the best RRH-Sector pair for each new connection demand by considering the user and operator objectives. This decision is based on an algorithm derived from the Markov decision process (MDP) with the objective of maximizing the integrated operator-user utility. In the second part of the framework, the load-balancing problem is addressed via optimization of the RRH-Sector-BBU dynamic mapping formulated as a linear integer-based constrained optimization problem. We compare solutions for this problem obtained by several evolutionary algorithms such as: Bee colony (BCO), Cuckoo search (CUCO), Genetic algorithms (GA), and Particle swarm (PSO). Finally, we evaluate the proposed solutions using extensive simulations. The results show that the proposed RRH-Sector selection scheme provides significant gains in terms of the operator reward and the connection blocking probability when compared to the received signal strength (RSS) method. Furthermore, the evolutionary algorithms are compared with the exhaustive search method that gives the RRH-Sector-BBU optimal mapping. The results show that, in most of the considered scenarios, the proposed algorithms reach the optimal solutions in terms of the number of blocked users, number of handovers, and BBU power consumption. Therefore, the proposed framework enhances the QoS and optimizes the network performance that balances the load across the network.

Keywords: Cloud radio access network, load balancing, Markov decision process, QoS.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AC	Admission control
ARPU	Average Revenue per User
BCO	Bee colony
BS	Base stations
C-RAN	Cloud radio access network
CIO	Cell Individual Offset
CoMP	Cooperative multi-point processing technology
CUCO	Cuckoo search
CQI	Channel Quality Indicator
CSI	Channel Status Information
ELR	Egg-laying radius
eNB	Evolved Node B
ES	Exhaustive search
FD-RR	Frequency Domain Round Robin
GA	Genetic algorithms
GPOS	Giga operation per second
GSM	Global System for Mobile Communications
HO	Handover
ICI	Inter-cell interference
KPI	Key performance indicator
MCS	Modulation coding scheme
MDP	Markov decision process
MME	Mobility management entity
NP	Network performance
PDCCHs	Physical Downlink Control Channels
PMI	Pre-coding Matrix Indicator
PtP	Point to point transceivers
PSO	Particle swarm

PSS	Primary synchronization signals
PUSCH	Physical Uplink Shared Channel
PUCCH	Physical Uplink Control Channel
QoE	Quality of experience
QoS	Quality of service
RAN	Radio Access Network
RBs	Resource blocks
RF	Radio frequency signals
RI	Rank Indicator
RRH	Remote Ratio Head
RSS	Received signal strength
SNR	Signal-to-noise ratio
SINR	Signal-to-interference-noise Ratio
SON	Self-organizing network
SSS	Secondary synchronization signals
TDD	Time-division duplex
UE	User equipment
WCDMA/TD-SCDMA	Wideband code division multiple access

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

$\bar{\lambda}_j$	Average arrival rate for accepted class *j* connections
$\bar{\lambda}_j^k$	Average arrival rate for accepted class *j* user connections in sector k
λ_j	Class *j* user arrival rate
λ_1, λ_2	Class 1 and class 2 users arrival rates
μ_j	Class *j* user connection rate
μ_1, μ_1	Class 1 and class 2 users connection rates
$\overline{c_j}^k$	Average cost rate for accepted class *j* user connections in sector k
$\overline{c_j}$	Average utility penalty rate for accepted class *j* connections
c_{RBG}	Price per unit bandwidth
$c_j^k(U_j^k)$	User utility penalty rate as a function of the class *j* user utility for sector k.
$g_j^k(x,\pi)$	Operator net-gain that can be obtained from accepting a class *j* user at sector k in state x
$g_j^k(z,\pi)$	Operator net-gain that can be obtained from accepting a class *j* user at sector k in state z
j	User class
$\hat{k}$, $\hat{k}_1$ and $\hat{k}_2$	Constants to express the penalty as a fraction of the reward parameter
k	Sector
$P_{R(\varepsilon,n,k)}$	Received signal power
$p_j^k(x,\pi)$	Shadow price for sector k and class *j* user when the sector is in state *s* under sector policy π
$p_j^k(z,\pi)$	Shadow price for sector k and class *j* user when the network is in state *z* under sector selection policy π
$\bar{R}(\pi)$	Average reward from the network under node selection policy π
$r_{\varepsilon,n,k}$	Data rate

r_j	Reward rate for class j user
T_1, T_2	Thresholds delimiting the user utility values zones
UM_i	Set of users served by RHH_i
$U_j^k(C, W)$	Aggregated user utility for sector k, class j user and for all criteria
$\mathcal{U}_j^k$	Aggregated utility for sector k, and class j user
$u_j^k(c_i)$	User utility of a class j user for sector k and for a given criterion c_i
$u(x)$	Utility value for a given criterion x
$V_q^s(x, \pi)$	Value function for node s in state x under node selection policy π at iteration q
w_o	Weight for operator utility
w_u	Weight for user utility
w_{pr}, w_{dr}, w_{rss}	Weights for the price, data rate, and RSS

INTRODUCTION

The internet of things (Tsagkaris et al.) era has changed the wireless access demands dramatically in recent years. The need to connect devices wirelessly caused an unwitnessed volume surge in mobile data traffic. By 2023, around 29.3 billion global mobile devices and connections are predicted to be active, and another prediction suggests reaching 50 billion (Cisco, 2016-2020). Consequently, mobile operators need to search for achievable solutions that can provide a balance between reaching a satisfactory level of service to customers and costs to optimize between profit and growth. It is evident that the radio access network (RAN) represents the most valuable asset for any mobile operator to achieve the Quality of service expected by the customers and yet achieve profitability. However, traditional RANs architecture is by far getting more expensive to keep mobile operators competing in the future. To grasp this dilemma, it's crucial to shed light on the main characteristics of the traditional RANs architecture. Most RANs consist of three main features which are as follows:

- **Connectivity**: Normally, there are a certain number of sector antennas that handle connectivity within a coverage area where transmission and reception are being processed. Each cluster of antennas is connected to a specific base station (BS).
- **System capacity**: Interference is a limiting factor that affects the spectrum capacity.
- **Location**: BSs need sites to be built on, and these platforms cost lots of money.

The above characteristics pose lots of challenges, especially for expanding over the future to meet the explosive growth in demand. There are two components that matter for mobile operators: the CAPEX and OPEX. Both expenditures will face a massive rise in expanding. This can be justified by the enormous initial investment that each operator will have to pay to accommodate more BSs, whether through building costs, site rental, operating costs, management support and so forth (Tschofenig et al., 2010). Moreover, the utilization rate can be considered low over the regular operation aside from peak times. On the other side, the processing power of each BS can be thought of as isolated. In other words, such power can not be shared between each BS and thus increasing costs and limit the spectrum capacity.

Furthermore, to provide Quality of service and offer broadband wireless services, the operators have to build more BSs and hence the power consumption rises significantly. It can be implied that the power consumptions increase proportionally with the number of base stations introduced. The increase in power consumption affects the OPEX and, above all, affects the environment (Zhou et al., 2003). It is essential to mention that the RAN equipment uses almost half of the power consumed by the BS; however, the other portion is used by the site support equipment like the air conditioning. Figure 1 shows the power consumption components based on China Mobile (Mobile, 2011), and it signifies the effect of RANs on the power consumption level. The challenge faced is visible, and the solution may be simple, which is reducing the number of BSs; however, relying on decreasing the number of BSs will affect the network coverage and decrease the capacity. Hence, there should be a solution to reduce power consumption without affecting the network capacity. Several technologies are available whether by providing a cleaner way of harvesting power such as wind, solar or other renewable energy sources or the use of smart control software to save power. This can be done through switching off idle carriers within the low traffic times; however, these solutions will only address the power portion used by the site support equipment. Such solutions will not solve the power consumption problem resulting from the rise in the number of BSs. Consequently, a change in the RAN infrastructure is the solution by adopting a technique that offers centralized BSs to reduce the BS equipment and the site support equipment while providing a mechanism for resource sharing and hence increase the utilization rate.

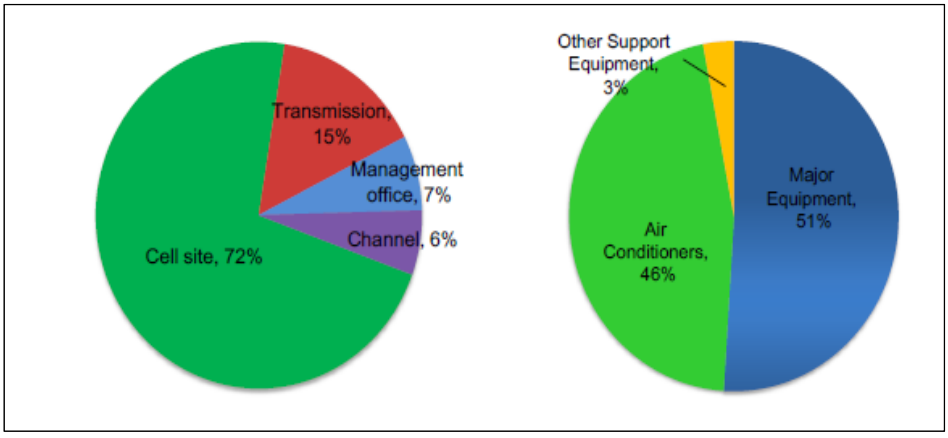

Figure 1 Power Consumption of Base Station

Not only this, but also the requirement that each mobile operator needs to include Global System for Mobile Communications (GSM), Wideband Code Division Multiple Access (WCDMA), and Long-Term Evolution (LTE) in their network requires constant upgrade aside from the operation cots. In brief, mobile operators will lack the flexibility in upgrading networks to accommodate the excess in demand in addition to the added services implied by the network future. Another term that puts the traditional RAN into a challenge is heterogeneous networks in which everything will be connected and managed by the network. This will increase the CAPEX and OPEX of such traditional architecture to be able to accommodate the new services, applications and support centralized interference management. Hence, it is an urgent need to find a convenient solution to enable flexibility and yet reduce costs. C-RAN came to be a viable solution to such a dilemma (Y. Lin et al., 2010). Figure 2 shows the difference between the traditional RAN and the C-RAN.

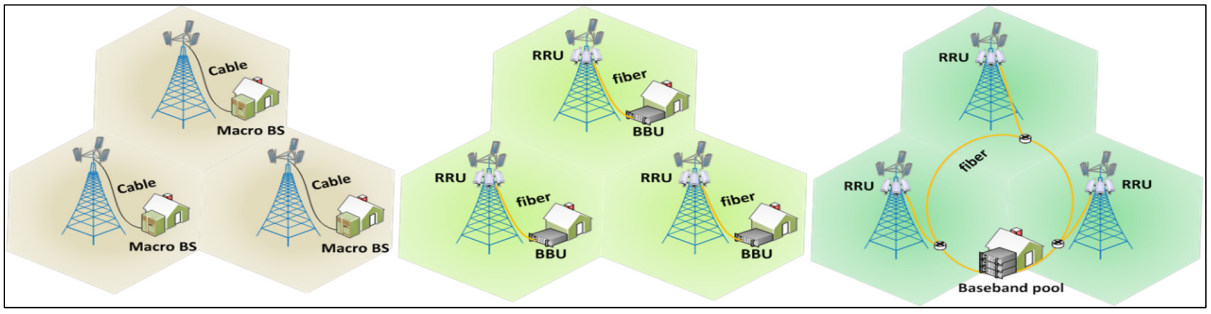

Figure 2 Different Network Architectures – Left: Traditional, Middle: Base station with RRH, Right: C-RAN Adapted from (Mobile, 2011)

The future RAN is required to reduce costs while obtaining high spectral efficiency. Besides, energy consumption should be diminished, and the flexibility to be able to accommodate different standards is indeed an essential requirement. To meet such needs, real-time C-RAN with the presence of a centralized base-band pool processing and cooperative radio with distributed antenna equipped by Remote Ratio Head (RRH). C-RAN will enable BS virtualization and allow for dynamic allocation and thus reducing the power consumption. Also, the utilization rate will be optimized, and thus increasing the throughput. The need for centralized pools is necessary to target the CAPEX and OPEX reduction since it will help in

reducing the number of site equipment to cover the same area. Moreover, the last technology, which is RRH equipped distributed antennas under cooperative radio, will meet the spectral efficiency requirement. C-RAN represents a way of delivery rather than a replacement to 3G/B3G standards. It provides a low cost and green way of deployment. There are other ways of RAN deployments rather than C-RAN. Each has its pros and cons; however, C-RAN can be considered a replacement for most typical RAN architectures such as Macrocell, Microcell, Pico cell and indoor coverage with the possible complementary deployment of other RANs for specific scenarios (Hwang et al., 2013). Keeping in mind the deployment of C-RAN with the aid of a centralized BBU pool and cooperative radio with RRH equipped distributed antennas; the C-RAN offers lots of advantages that can be summarized as follows:

- ***Expenditure reduction (CAPEX and OPEX saving):***
  - ***CAPEX****:* Adopting C-RAN will reduce the number of BS sites required for operation and providing the needed services. Virtualization is the key to such drastic change.
  - ***OPEX****:* The idea of centralized management and operation is the main advantage over typical RANs since the BBUs and site support devices can be added in fewer rooms and thus saving lots of OPEX, whether from the operating point of view or the maintenance point of view. The challenging aspect is that RRHs number will not be affected by deploying C-RAN; however, the advantage is that their functionality under the C-RAN umbrella will be simplified. In addition, their size and power consumption will decrease, and they will not require the same amount of site support and management as traditional RANs.

1. Motivation

The massive use of smartphones and applications for information and communications technologies is producing a huge growth in the demand for mobile broadband services with higher data rates and Quality of service (QoS). As stated by the Cisco Visual Networking, By 2023, around 29.3 billion global mobile devices and connections are predicted to be active, and another prediction suggests reaching 50 billion (Cisco, 2019). Consequently, the mobile

network operators must consider significant actions to adjust this enormous traffic growth. C-RAN is one of the promising solutions that offers a centralized network that can accommodate the expected high traffic demand. Hence, it is considered as one of the essential architectures that will define Mobile Networks (i.e. 5G) (Cisco, 2016-2020). Real-time C-RAN, with the presence of a centralized base-band processing pool and cooperative radio with (RRH) equipped with distributed antennas, is predicted to be a vital solution to face the challenges mentioned above (Alliance, 2013). This technology has many advantages that can be summarized as follows:

- The C-RAN is a cost-effective, low power radio-access points with small coverage service areas varying from tens to hundred meters connected to a centralized BBU (Forum, 2014).
- C-RAN can improve indoor/outdoor coverage and network capacity by reducing path loss and reusing the whole spectrum.
- C-RAN provides low operational cost due to its self-organization feature that self-configures, self-optimizes, and self-heals itself automatically without human intervention.
- C-RAN network gives a friendly network with a low transmission power environment.
- C-RAN saves the power of the RRHs and the life of UEs batteries.
- Better performance can be provided by balancing the load of the network.

As a result, providing QoS in this technology is a vital topic that needs to be addressed. It is proved that the achievable data rate of any terminal depends on the users to RRH-Sector selection decisions. That motivates us to study the user RRH-Sector selection to assure higher data rates and less blocking rates.

Moreover, a varying traffic environment can cause resources underutilization, and QoS as well as NP deterioration that is still a big challenge in C-RAN. This is because the overloaded RRHs-Sectors can face resource shortage that degrades NP when UEs try to connect to those RRHs-Sectors, although there are nearby under-loaded RRHs-Sectors that could serve these UEs. Therefore, this motivates us to find appropriate management and configuration methods

to maximize the NP of the C-RAN. This can be done through optimal RRH-Sector selection and dynamic allocation of BBUs resources to RRHs-Sectors.

2. Problem Statement

Despite the advantages that C-RAN offers, there are still some challenges to be addressed. The most important ones are Quality-of-Service (QoS) support and resource management. Some issues have been widely studied in the literature. However, many aspects have not been considered in those studies. The problems that need to be addressed for C-RAN are the following:

- **Service unavailability:** caused by the blocking of connections due to lack of resources or lack of network coverage (dead zones). Service providers have tried to correct the problem by expanding their network coverage and increasing the ability of cells.
- **Service interruption:** RRH-Sector pair selection should be optimized based on the user and operator utilities. The commonly used RRH-Sector selection method based solely on the measurement of the Received Signal Strength (RSS) does not guarantee user satisfaction since this satisfaction is usually based on several QoS metrics (data rate, delay, packet loss, etc.,). The weaknesses of this method are described in (Hashim et al., 2013), where the authors indicate possible underutilization of network resources, lack of fairness, and inappropriate load balancing in the network. The RSS metric is conventionally used for selection and handover (Benaatou et al., 2017; El Fachtali et al., 2016), but some authors (Mahardhika et al., 2015; Miyim et al., 2013) argue that it is only convenient to start the handover and then adding some other QoS criteria is needed for better results. Therefore, to maximize the utilization in the C-RAN network, a decision for RRH-Sector selection should consider the requirements and constraints of the user and the operator. From the user perspective, the application related QoS metrics and other criteria, for instance, the power consumption and the cost of the service of the device are of importance. Specifically, data rate and delay are essential for real-time applications, while low packet loss is essential for elastic traffic applications (web browsing, file transfer).

- **Imbalance and network overload:** a changing traffic environment can cause resource underutilization and network performance degradation that is still a big challenge in C-RAN (Ahmed et al., 2016). This is when the load distribution among RRHs-Sectors is not uniform. Consequently, the overloaded RRHs-Sectors can face resource shortage when UEs try to connect to those RRHs-Sectors, although there are under-loaded RRHs-Sectors neighbors that could serve these UEs. Therefore, proper management and configuration methods should be introduced to optimize the network performance of the C-RAN.
- **Illustrative example:** Figure 3 visualizes the C-RAN architecture. We are going to show the challenges with an example. The first challenge is RRH-Sector selection for each UE. We assume that each BBU serves three sectors, and each sector is defined with a specific color and consists of one or multiple RRHs. For example, let us assume that each sector can serve up to five connections coming from different RRHs that are being associated with that specific sector. Let us assume that UE (c) arrives at the network and wants to be served. Then due to its location, the network will select either RRH1-Sector6 or RRH4-Sector1 for him. Let us assume that the network selects RRH4-Sector 1 regardless of the state of each sector. Then UE (b) arrives and wants to be served. Note that UE (b) can only connect to RRH4-Sector 1, and this sector reached its maximum capacity (i.e., five connections come from RRH2 and RRH4). Thus, this connection is going to be rejected in RRH4-Sector 1 due to the not optimal RRH-Sector selection for UE (c). However, this rejection could be avoided if UE (c) is allocated to RRH1-Sector 6.

  The second challenge is balancing the load among the Sectors. It is clear from Figure 3 that sector 4 has five connections that come from RRH 7 and RRH 8, so any new connection that comes to that sector will be rejected. On the other hand, sector 3 has only one connection that comes from RRH 5. Since the C-RAN has the dynamic ability to reconfigure the logical connections between RRHs and Sectors, RRH7 could be switched to be served by sector 3 instead of sector 4. In this case, the load is going to be balanced among sectors, and each considered sector can accept two new connections.

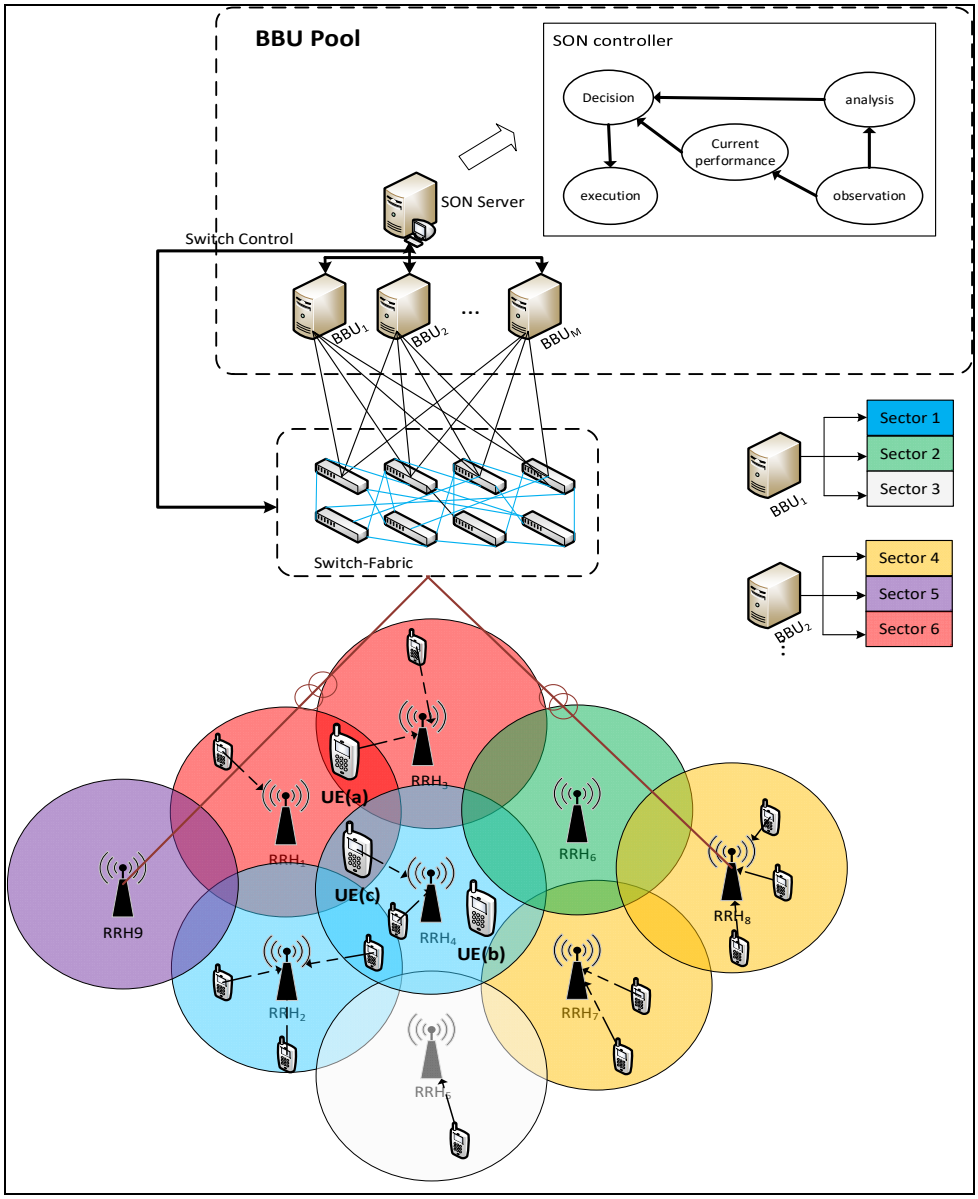

Figure 3 C-RAN architecture illustration

3. Research Objectives

The main objective of this thesis is to develop an RRH-Sector selection and load-balancing, among the sectors and BBUs, framework that aims at the optimization of the NP, QoS and operator revenue in C-RAN. The RRH-Sector selection maximizes the joint utility which includes the user and operator utilities. On the other side, the load balancing aims to optimize

the NP in the C-RAN. This is done based on various Key performance indicators (KPIs) such as the standard deviation of the sectors load, number of forced handover blocked users, number of handovers, and the power consumption. Each KPI has a defined weight that is chosen by the network operator according to its preferences.
Generally, chapter 6 could be more explicit about how it demonstrates these problems have been solved and should answer the following questions:

- How to maximize the number of UEs admitted into the C-RAN;
- How to maximize user QoS in the C-RAN;
- How to maximize total operator reward in the C-RAN;
- How to maximize network performance (NP) in the C-RAN; and
- How to improve the load balancing in the C-RAN.

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

4. Methodology Overview

For the RRH-Sector selection we propose a model based on the integration of the operator utility and the user utility. This model can be also considered as a short-term load balancing method between sectors since the users located at the boundary of two sectors can be allocated to either of them. The operator objective is to maximize the utility defined as the mean value of reward from the network being a sum of rewards from admitted connections. The maximization of the operator utility is based on sector shadow price and node net-gain concepts derived from Markov Decision Process (MDP) theory. Then, we introduce user utility, which is the satisfaction of the user with respect to several QoS criteria. The user utility is calculated using the sigmoid function taking into account several criteria (e.g., price, data rate, delay, etc.). It is used to calculate the user utility penalty that is used to reduce the operator utility by integration of the user utility penalty within the Markov Decision Process that optimizes the average operator reward which represents the operator utility.

The load balancing among sectors and among BBUs is optimized by finding the best RRH-Sector-BBU mapping. The objective of this optimization is formulated using various NP KPIs such as the standard deviation of the sectors load, number of forced handovers blocked users, number of handovers, and the power consumption. Each KPI has a defined weight that is chosen by the network operator according to its preferences. To solve this optimization problem with unknown structure, very little information, and limited computational capacity, we selected four evolutionary algorithms: Bee colony (BCO), Cuckoo search (CUCO), Genetic algorithms (GA), and Particle swarm (PSO).
Finally, we use MATLAB software to simulate selected scenarios in order to evaluate the proposed framework and its performance.

5. Thesis Contribution

There are three main contributions of the presented work. The first one is using the MDP for the RRH-Sector selection model that allows the dynamic nature of user activities as well as variations in data transmission conditions to be considered. We use a decomposed form of MDP with the concepts of shadow price and net gains applied to reduce the complexity of calculations due to a large number of states that allows its implementation in practical networks.

The second contribution consists of designing a simpler utility-based RRH-Sector selection model that considers both the user utility and the operator utility jointly. The main reason for selecting the utility functions is that it indicates the level of satisfaction for any metric through a numerical value. Thus, it is simpler to assess these numerical values at the same time to achieve the optimum decision throughout the selection process compared to the heuristic algorithms used in (Fedrizzi et al., 2016).

The third contribution is introducing a power-efficient load balancing algorithm through dynamic RRH-Sector-BBU mapping in the C-RAN. In contrast to the existing works which

did not consider the power consumption during the load balancing process, our algorithm balances the load in the C-RAN with the minimum required power to operate the network by switching off the underutilized BBUs. The RRH-Sector-BBU load balancing optimization objective is formulated using various NP KPIs such as sector load standard deviation, number of blocked users, number of handovers, and the power consumption.

6. Thesis Outline

The chapters of the thesis are structured as follows:

Chapter 1 (LITERATURE REVIEW) introduces information about C-RAN architectures, resource management, and the UE and RRH selection process from the literature. Also, the related work on load balancing techniques proposed by other researchers is presented.

Chapter 2 (C-RAN SYSTEM MODEL) introduces the C-RAN system model when LTE is applied. All the necessary parameters and assumptions to operate the C-RAN are presented.

Chapter 3 (RRH-SECTOR SELECTION MODEL BASED ON MDP WITH PENALTY) presents the proposed RRH-sector selection in the downlink LTE C-RAN network. The model employs the user utility as well as the operator utility that jointly take users' QoS demands and network profit into account. The proposed model uses the concepts of RRH-Sector shadow prices and RRH-Sector net gains derived from MDP decomposition with the objective of maximizing the integrated operator-user utility.

Chapter 4 (DYNAMIC RRH- SECTOR-BBU MAPPING AND RELATED KPIs) introduces development of the load-balancing algorithm. The algorithm balances the load among the sectors and BBUs. Its objective is to optimize the NP in the C-RAN, which is done based on various KPIs. Those KPIs are the standard deviation of the sectors load, number of forced

handover blocked users, number of handovers, and the power consumption. Each KPI has a defined weight that is chosen by the network operator according to its preferences.

Chapter 5 (EVOLUTIONARY ALGORITHMS FOR RRH-SECTOR-BBU MAPPING) presents the selected four evolutionary algorithms: Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Bee Colony Optimization (BCO) and Cuckoo search (CUCO). All of them belong to the group of swarm-based optimization algorithms and are used to find the optimum RRH-Sector-BBU mapping to balance the load across the network.

Chapter 6 (PERFORMANCE EVALUATION OF THE PROPOSED MODELS) presents the simulation results for the RRH-Sector selection and load balancing framework using MATLAB software. The simulation results for the RRH-Sector algorithm are compared with commonly used network selection techniques based on the received signal strength (RSS). The metrics used to evaluate the performance are the operator reward, the users blocking probabilities and the users' average data rates. Also in this chapter, the numerical results for the proposed power efficient load balancing approach, using the BCO, PSO, CUCO, and GA algorithms, are presented. These results are compared with the exhaustive search method, which finds the optimal RRH-Sector-BBU mapping.

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

The first part of this chapter describes in more details possible C-RAN architectures including the one treated in this thesis. Then some background concepts regarding the different existing resource management techniques, and the literature review on RRH-Sector selection and load balancing in C-RAN are introduced as well.

1.2 C-RAN architecture

C-RAN is going to provide a drastic amount of saving on the operation, management and site rental. Moreover, as an extra advantage, the RRH will only require the installation of auxiliary antenna feeder systems and thus enabling a speed up in network establishment.

- **Green Infrastructure and Energy efficient:**
  - **Site reduction**: It is acknowledged that the number of sites in C-RAN deployment will be reduced. Moreover, the C-RAN is alleged to be an eco-friendly infrastructure for this reason. BSs sites reduction through centralized operation will offer a decrease in using power-based equipment such as air conditioning and other site support devices and hence reduce the consumed power (CLI et al., 2014).
  - **Interference reduction**: The cooperative radio with RRHs will help to reduce the interference and thus allow for a reduced distance between the RRHs and EUs and hence higher density of RRHs (Hoydis et al., 2011).
  - **Power consumption reduction**: The power required for signal transmission will be lowered, which will have an impact on the power consumption of the RAN while giving a longer battery lifetime for the UE. This will allow for utilizing smaller cells while not affecting the network coverage quality. Besides, virtualization is a fundamental aspect of power consumption. In C-RAN, the BBU pool can be thought of as a shared resource

among the virtual BSs and hence optimization through resource allocation can be achieved. Consequently, the utilization rate can be enhanced while lowering the power consumption requirements. For example, when a virtual BS went idle at night, the processing power can be turned off or reduced without affecting its capability of performing the required services (Mobile, 2011).

- **Capacity Enhancement:**
  - This can be achieved through joint processing and scheduling, where the effect of inter-cell interference (ICI) will be reduced and thus enhancing the spectral efficiency. In fact, virtualization assisted in forming a framework for virtual BSs to work under where they can share a large physical BBU pool allowing for signalling, traffic data and channel state information (CSI) of active UE's sharing across the system (Hoydis et al., 2011).
  - Such enhancements allow for some technologies to operate in conjunction with C-RAN, such as cooperative multi-point processing technology (CoMP in LTE-Advanced).
- **Traffic management:**
  - **Non-uniform traffic handling**: C-RAN can handle non-uniform traffic load through load-balancing functionality in the distributed BBU pool. Usually, the BBU pool offers broader coverage than the traditional BS; hence, the non-uniform traffic from UEs will still be within the same BBU pool and thus can be distributed within a virtual BS. Moreover, the increased radius of the BBU pool will allow serving the RRH despite its dynamic change due to the movement of the UEs.
  - **Smart Internet Traffic offload**: This intelligent technology provides a gateway to offload the core network of operators from the ever-increasing internet traffic from smartphones and portable devices. This will be advantageous in many aspects since it will reduce the core network traffic, decrease the latency for the users and providing them with a better experience, decrease the back-haul traffic and cost and thus enhancing the core network operation.

Although C-RAN is regarded as a solution for several challenges based on the advantages stated above, however, the C-RAN must optimize the performance metrics that define the QoS as following:

- Data rate: the number of packets received per time unit represents the data rate, which might be normalized by dividing the rates of the received packet over the sent packet.
- Blocking probability: it is defined as the probability that some UEs cannot achieve the minimum required data rate in the network. The objective is to minimize the blocking probability as much as possible.
- The number of handovers: handovers may cause packet losses and delays and consequently lead to a poor UE's Quality of experience (QoE). The objective is to minimize the number of handovers as much as possible when realizing a balanced-load network.
- Standard deviation: The range of Standard deviation is in the interval [0, 1], with a smaller value indicating an extremely load balanced distribution between all active base stations. Therefore, standard deviation minimization is one of the objectives to reach a very well-load balanced network.

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

- **C-RAN architecture based on RRH and BBU functions**

Based on the different splitting functions between RRH and BBU, there are two types of C-RAN architectures: the first solution is named "full centralization," where the BBU integrates layer 1 (base-band), layer two and layer three functions.

This type has several benefits as it can be upgraded easily, and it can increase network capacity; it also supports multi-standard operation and sharing of the resources. Also, it is more suitable for multi-cell collaborative signal processing. Its drawback that it needs high bandwidth between the BBU and to carry base-band signal.

The second solution is named" partial centralization," where the RRH has the base station functions besides the radio functions, while the BBU still having the other higher layer function. The advantage of this type is "small centralization," with partial BBU functions

centralized into a single central point that is linked with the remained remote BBU via dark fibre. With such design, the central point can control the wireless resources in every cell on a global level and even recognize the joint transmission or joint reception on the Physical layer to enhance cell edge performance. The main drawback of this architecture is that it still needs remote equipment rooms. One-body type base station is not desired from the viewpoint of system management and future upgrade. Moreover, the system improvement can be affected due to the delay in the information exchange (Y. Lin et al., 2010).

As shown in Figure 1.1 below the different function splitting scheme:

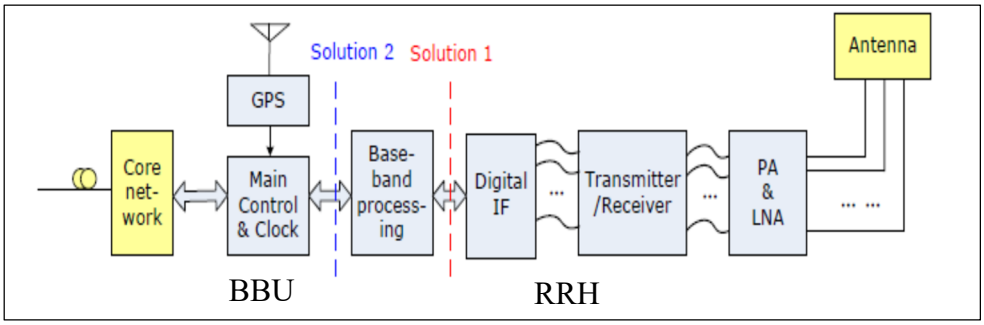

Figure 1.1 C-RAN architecture based on RRH and BBU functions

Brief description of the function of each element on the C-RAN system:

- **RRH (Radio resource head)**: it transmits and receives wireless signals, provides the interface to fibre connections, and it changes the digital base-band signals that sent from the BBU into radio frequency signals (RF). Then it amplifies the power of the signal and sends it to the UE. On the opposite side, when UE sends RF signals to the RRH, those signals are amplified and transformed into digital base-band signals, and then they send to the BBU. Also, each RRH is equipped with a specific number of antennas that may be different from one to another.
- **Buffer**: it is a physical memory storage area, and its function is to temporarily store data when it is being transferred from the RRHs to the mobile user.
- **BBU (Baseband units)**: works as virtual base stations for base-band signal processing and network resource allocation optimization, as it provides connectivity to RRHs and core network, where it is responsible for digital base-band signal processing, monitoring control

processing. When the core network sends IP packets, those packets are modulated into digital base-band signals and then transferred to the RRHs. Still, when the RRHs send digital base-band signals, those signals are demodulated, and IP packets are sent to the core network.

- **Fronthaul links**: they are the fibre cables that connect the RRHs with the BBUs.
- **Power consumption:**
  - In case several RRHs are used to transmit data, the power participated for a cooperative transmission of a shared stream from several RRHs to specific a UE can be represented by two factors. The first one is the total power to transmit the shared stream to this particular user, and the other element represents the contribution of this specific RRH.
  - In case only one RRH is used to transmit data: the power used by a specific RRH is used to send data to a particular mobile user.
- **Total power**: is the summation of the power used to transmit the data through the first case, which uses several RRHs at the same time to transfer the data, plus the summation of the power used in the second case to transmit data to a mobile user through only one RRH.
- **Load balancer**: it flexibly manages the admission to the shared radio resources based on the NP parameters, and it determines some or all of these functions:
  - Determining which UE should be served in each time interval of the transmission.
  - Allocating the physical resource blocks for each UE.
  - Determining which coding and modulation type should be used during the transmission for each UE.
  - Combining signals from and divide signals to the RRHs by linking the electrical and logical signals between the servers and RRHs.
  - Regaining and synchronizing the clocks between the RRHs.
  - Working as a “smart router” between the servers to balances and distribute the load.
  - Reducing the latency on the system (offload time crucial control functions from servers).
  - Processing signals received from multiple RRHs.
- **Scheduler**: Map the packets to the BBU and take the decision about whether to accept the packets or not.

- **The topology of the connection in C-RAN**

The topology of the connections in the C-RAN system depends on the number of available fibre cables installed. Therefore C-RAN can be constructed for macro networks with either a star or tree topology if the operators have numerous access fibre resources and enough aggregative level fibre resources. Note that the reliability of the network can be enhanced when a star topology is used. It is usually appropriate for single-radio access technology cloud and multi-radio access technology cloud. On the other, when fibre resources are limited, fibre ring topology is used for single- radio access technology sites. For areas without any fibre resources, microwave transmission could be an alternative (Y. Lin et al., 2010). Figure 1.2 below represents fibre topologies for C-RAN; the first one is the star, while the other is ring topology.

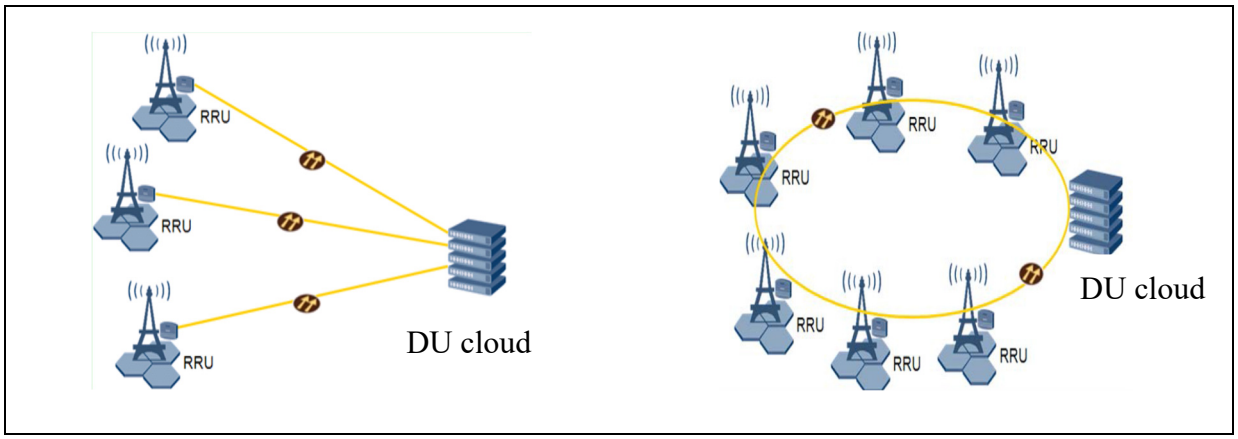

Figure 1.2 Left: C-RAN star topology, Right: C-RAN ring topology (Mobile, 2011)

- **Types of connections between BBU and RRH**

There are two types of connections between BBU and RRH, the first one named as distributed while the second one is centralized, the type is chosen relying on different resource management in the BBU pool.

For the first type, which is the distributed type, each RRH is directly connected to a dedicated BBU. This type is realized easily and simple, also it is named as “one to one” mapping. It needs several BBUs to be operating and generating frames that use a lot of power in the BBU pool.

It is not useful to achieve the benefits of the joint signal processing and C-RAN central controlling. The second one, which is a centralized type, all the RRHs could be connected to a central device or a switch, which can manage the processing resources in the BBU pool either for a group of RRHs or an individual RRH. This type has many benefits in terms of the sharing of resources and power efficiency using joint scheduling. In addition to that, the execution of efficient interference prevention and avoidance methods within multiple cells could be implemented in this type. Also, it offers the capability to manage the operation of RRHs, for example, turn RRHs on and off depending on the traffic variations in different cases.
The front haul connections and links are recognized by different technologies, such as wireless or fibre, and classified into two types, the first one is ideal without bandwidth constraints, and the second one is non-ideal (Y. Lin et al., 2010).

Optical fibre is considered an ideal front haul for C-RAN since it offers a high data rate and large bandwidth. For example, the performance specifications for the "next-generation passive optical network 2" is indicating that it has 40 GHz bandwidth for downstream and 10 Gbps for upstream, and can extend up to 40 km. "Very-high-bit-rate digital subscriber line 2" uses up to 30 MHz of bandwidth to deliver 100 Mbps speed for both upstream and downstream within range of 300 to 400 m. Also, wireless backhaul can be deployed as they are less expensive and quicker than fibre. They generally work on the licenced band using ruses methods such as a relay. They could also utilize microwave technology with a carrier frequency range between 5 and 40 GHz. However, because those frequencies have limited available bandwidth, data rates of only a hundred Mbps can be provided (Mobile, 2011).

- **Selected C-RAN architecture**

In this thesis, we are going to study the C-RAN architecture where the type of connections between BBU and RRH is star, the link between them is going to be optical fiber cable, and the selected C-RAN architecture follows "full centralization" solution. Moreover, each group of RRHs are forming a sector where this concept helps to increase the capacity in the C-RAN. As shown in Figure 3 each BBU serves three sectors, and each sector is defined with a specific color and consists of one or multiple RRHs.

1.3 Related work on resource management and load balancing

In this section, we will present some existing techniques in the literature for resource management for C-RAN.

1.3.1 Resource management

The resources in the wireless systems are represented by the channels in terms of frequency, time intervals (time slots), transmission power and battery power represent. Proper resource management can help service providers reduce costs, increase revenues while ensuring a better QoS. Resource management can improve outcomes in terms of blocking probability and ensure continuity of service. An overview of C-RAN, its advancements, initial prototypes and field trials, and overview of practical deployment of C-RAN can be found in (Checko et al., 2015; Mugen Peng et al., 2016). The works of (He et al., 2016; Simeone et al., 2016; C. Yang et al., 2015) provided a research initiative to establish Software Defined Network (SDN) and Function Virtualization (NFV) Software for C-RAN.

The main issue which mainly affects the handover is admission control, bandwidth reservation, and the scheduling policy (Van Quang et al., 2010). Figure 1.3 shows these important components.

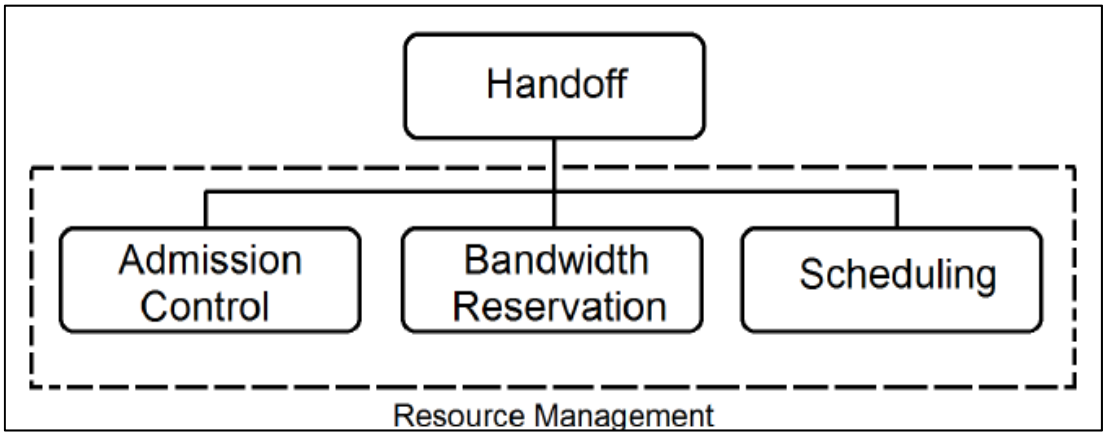

Figure 1.3 Resource management in connection with the handover

More comprehensive reviews on C-RAN from the resource management point of view, physical layer and the challenges faced by the network, alongside possible solutions, can be found in (Quek et al., 2017; Tohidi et al., 2020; Venkataraman et al., 2017).

1.3.1.1 Admission Control

The admission control (AC) creates the conditions of acceptance of the various call categories in the network to guarantee a low rate of blocking and an optimal call acceptance. Therefore, the purpose of admission control is to determine the possibility of admitting a call in the network without affecting the required QoS by the call context. A call is banned if the above condition is not fulfilled.

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

Other policies based on revenue have been proposed to maximize the new call acceptance, which means maximizing the revenue since each new call is a potential source of income. In (Nelakuditi et al., 1999), a resource allocation strategy was presented to minimize the call blocking rate when receiving a new call in a saturated cell (with no available resources). The

strategy is to move a current call to a neighbouring cell to release resources currently used and reallocate it to the new demand. In the LTE standard (Access, 2009), the user association algorithm depends only on the Received Signal Strength Metric (RSS). Though, that algorithm was not intended for a heterogeneous network. In (NTT, 2010), a bias is included to the Reference Signal Received Power (RSRP) to expand the small cell coverage. In (Qualcomm, 2010), the user is going to select the base station which ensures the minimum path loss values. In (Saad et al., 2014), based on matching game theory, the uplink user association method is introduced to solve the association problem for small cell networks. Though, this method considers the fixed small cell networks and only the access link between users and various eNBs.

More specifically, in C-RAN, the issue of the selection between UEs and RRHs is under extensive research. The authors in (You et al., 2019) aimed to improve the UEs' capacity performance through a framework that mutually optimizes UEs to RRHs association (i.e. CoMP selection) as well as resource allocation in time-frequency, in terms of user-centric demand scaling.

Moreover, the authors in (C. Pan et al., 2017) targeted a User-centric C-RAN in a MIMO layout. They aimed to minimize the possible power consumption within the network through solving a precoding and RRH optimization challenge jointly.

Furthermore, the authors in (Wang et al., 2017) targeted the RRH-UE clustering through solving a beamforming vector optimization problem. In (M.-M. Zhao et al., 2019), the goal was to achieve a balance between the minimization of the transmission power and the reduction of the front haul traffic by mutually optimizing the downlink beamforming vector, user association, and the caching placement for every RRH.

On the other hand, in (Kaiwei et al., 2017), they used the weighted minimum mean square error (WMMSE) method and studied the BBU scheduling to provide an energy-efficient approach. In (L. Liu et al., 2017), the authors worked on a cross-layer framework to enhance the

throughput of the network by jointly optimizing the network and physical layer resources. Moreover, they provided an overall design that is optimized from different aspects such as user RRH association, RRHs beamforming vectors and network coding-based routing.

1.3.1.2 Bandwidth Reservation

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

1.3.1.3 Scheduling

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

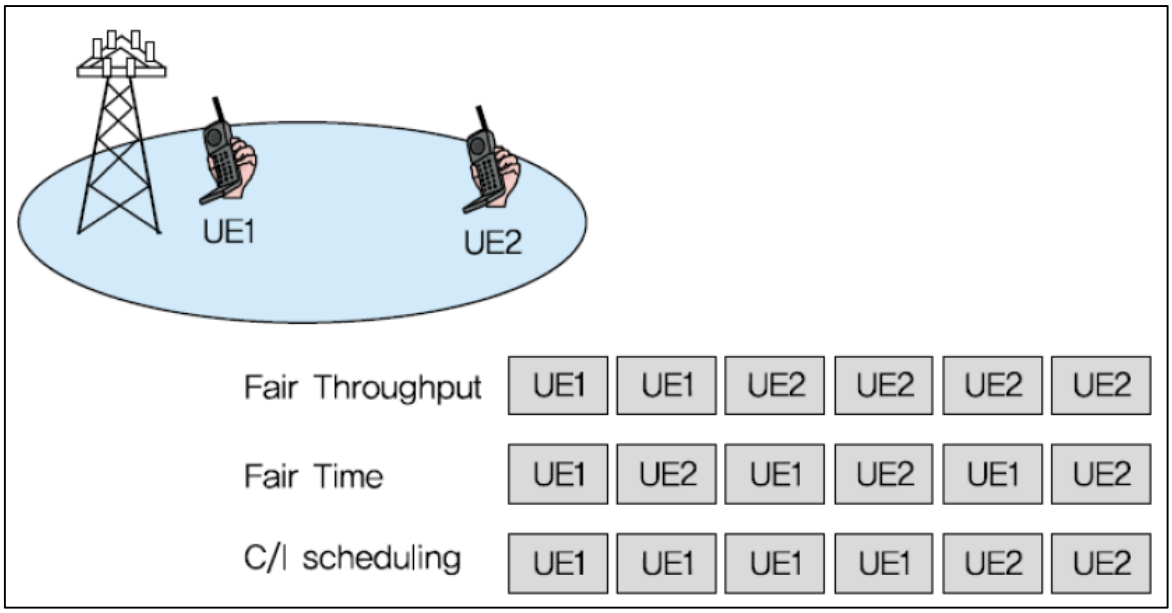

Figure 1.4 Three Conventional Scheduling Algorithms
Taken from NSN: LTE MAC/RLC/PDCP/RRC (2011)

The best-known scheduling algorithms in the literature depicted in Figure 1.4 are as follows (Dahlman et al., 2013):

- Round Robin Algorithm (also known as Fair Time scheduler): also called “Fair Time scheduler” that share resources equitably among UEs without considering the radio channel state.
- C/I scheduler: this one considers the radio channel state and primarily seeks to maximize the radio resource efficiency without considering fairness among UEs.
- The Fair Throughput algorithm: it provides a fair rate for all UEs even if the resource usage is far from optimal.

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

1.3.2 Handover Management

The main objective of Handover Management in C-RAN is to assure the continuity of the service during the handover process. Handover is the most sensitive point in the convergence of any two adjacent RRHs-Sectors. This transition is assumed homogeneous and transparent to UEs, which implies that the mobile UE must be auto-configured with the new settings without user intervention. Figure 1.5 shows the types of handover: horizontal and vertical handovers.

- **Horizontal Handover**

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

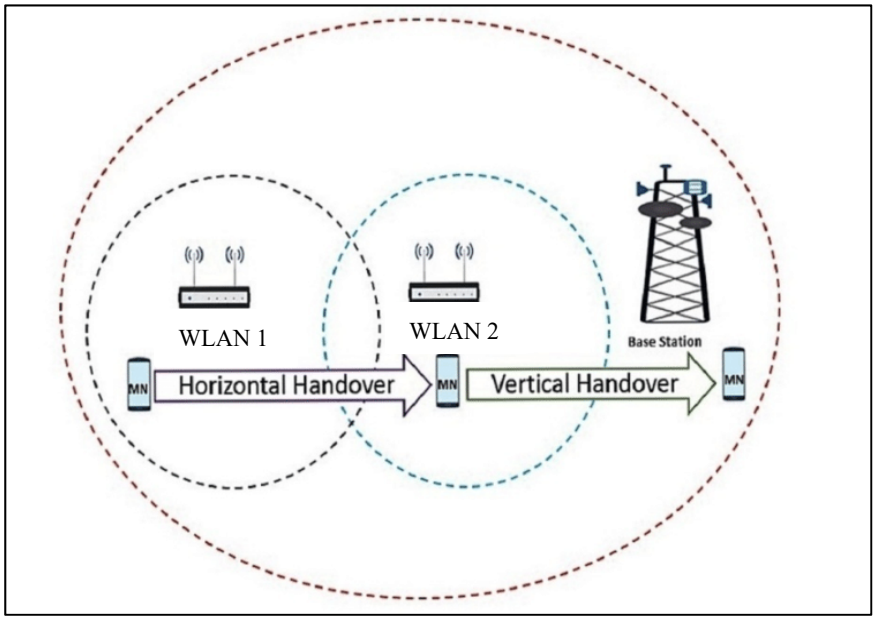

Figure 1.5 Horizontal and Vertical Handovers

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

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

Several Handover methods were presented to overcome the challenge associated with the frequent handover of mobile equipment. One of the optimized handover processes for the LTE network is designed based on the coordinated multiple point (ComP) transmission technology and dual-vehicle station coordination mechanisms (Luo et al., 2012).

- **Vertical Handover**

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

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

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

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

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

In fact, several ways of research have been developed only for the static scenario with a focus on maximizing throughput across both networks while minimizing the number of handovers.

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

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

Work in (Chou et al., 2006) used the signal-to-noise ratio (SNR) and traffic type as the metrics for the HO decision. Its goal was to maximize the throughput of the network and minimize the ping-pong effect. Lin, H et al. proposed a QoS-Based Vertical HO. Besides, reference (K. Yang et al., 2007) uses the combined effects mentioned above, including signal-to-interference-noise Ratio (SINR) to make HO decisions for multi-attribute QoS considerations. Still, all the above-

mentioned proposed techniques were studied from the core network point of view; however, integrating Wi-Fi in RAN makes it a different issue that needs to be investigated in terms of mobility and resource management.

1.3.3 Load Balancing

The network parameters can be self-optimized to the cell current load and in the neighbouring cells to enhance the capacity of the system. Also, human intervention is significantly minimized. The QoS experienced by UEs with load balancing should not be worse than that without load balancing. The main objective is to deal with the uneven traffic load by optimizing the network parameters. Nonetheless, the number of handovers required to do the load balancing should be kept as minimum as possible.

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

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

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

Researchers in (Huang et al., 2015), a multi-traffic load balance (MTLB) scheme is introduced for traffic load balancing and improving the capacity of the network with a proper handover technique. A new cell selection is adapted to improve the UEs Quality of service. Moreover, the TTT (time to triggering) and handover threshold is adaptively changed to decrease the rate of call drop with a more balanced-load network. Two conditions are accounted for the handover procedure: The signal strength condition and the RB condition. That helps in avoiding the wrong eNB or unnecessary handovers.

Not only UEs mobility could be used in balancing the load among the network cells, but also various types such as coverage and capacity optimization could be considered (Yamamoto et al., 2012). Once a small cell is discovered to be overloaded, the self-organizing network (SON) owns a role that reduces the power and consequently forces some edge-UEs to be offloaded to the lightly loaded cells of the network.

In regard to the load balancing in the C-RAN, the mapping between BBU and RRH is not highly considered in research. Some of the limited work on this issue can be found in (Y.-S. Chen et al., 2018; K. Lin et al., 2017; Namba et al., 2012; Sundaresan et al., 2016; Z. Yu et al., 2016). A mapping technique between the BBU and RRH, which is dynamic in nature, has been proposed by the authors of (Y.-S. Chen et al., 2018). They used a borrow-and-lend approach to provide a network load balancing while improving the throughput. The approach switched the RRHs from the overloaded BBUs to the underutilized ones.

In (Namba et al., 2012), they aimed at adjusting the BBU-RRH configuration using an adaptive switching and semi-static methods. They considered the traffic loads at peak hours for the RRHs in the system through a specific time interval.

The need to accommodate user traffic profiles dynamically led the authors of (Sundaresan et al., 2016) to propose a scalable algorithm that relied on ideal transmission methods through reconfiguring BBUs and RRHs.

In (Z. Yu et al., 2016), the authors worked on TDD-based heterogeneous C-RAN. They studied its energy-saving capabilities and its traffic adaptation and tried to enhance that by modifying the BBU-RRH logical connections.

The authors in (K. Lin et al., 2017) worked on the clustering problem where they proposed a spectrum allocation genetic algorithm (SAGA) to enhance the NP of the network through efficiently utilizing the available resources.

In (L. Chen et al., 2020), the authors proposed a deep-learning-based Multivariate Long Short-Term Memory (MuLSTM) method to gather the space-time types of mobility and traffic for precise estimation. Also, they defined the RRH-BBU mapping as a set separating problem and introduced a Resource-Constrained Label-Propagation (RCLP) algorithm as a solution.

The authors in (Han et al., 2019) considered the RRH-BBU Mapping in C-RAN for high-speed railway (HSR) scenarios. They used the graph theory as they abstracted the C-RAN into a graph so they could allocate the resources dynamically.

The clustering problem was addressed in (X. Chen et al., 2014), where a dynamic algorithm was proposed based on a greedy multi-objective optimization scheme, which is non-linear. This optimization problem was solved using the scalarization technique alongside a linear model. In this paper, the RRHs capacity was increasing in joint conjunction with enhancing the energy-efficient through utilizing this dynamic clustering technique in the downlink.

In the same line of work, the authors of (Shi et al., 2015) proposed a greedy algorithm to minimize the power consumption used by RRHs and the transport network. They used on and

off switching mechanism to enable joint selection of RRHs and thus minimize the power consumption. Furthermore, they utilized two group sparse beamforming methods, which are the bisection and iterative, to turn the RRHs off.

The idea of self-organizing C-RAN was exploited in (Khan et al., 2017), where the authors worked on decreasing the number of blockages to improve the QoS of the users. They relied on a third-party server called “Host Manager” that was located inside the BBU pool to modify the resource mapping between BBUs and RRHs dynamically. This server tries to select the best configurations using a genetic algorithm to minimize the load conditions.

The authors of (Mishra et al., 2016) relied on the well-known First Fit Decreasing (FFD) algorithm to provide a lightweight load-aware technique that can meet the RRHS computational resource demands with a reduced number of active BBUs. They followed the many-to-one mapping approach as a mapping scheme between the RRHs and BBUs.

The work of (Q. Liu et al., 2016; M. Peng et al., 2015) studied the energy-efficient resource assignment allocation in heterogeneous C-RANs. Finally, the work of (Dhifallah et al., 2015) aimed at reducing power consumption through a two-stage iterative heuristic algorithm by considering the hybrid backhaul, which relies on wireline and wireless connections. The first stage works on optimizing the selection of the BS and the beamforming jointly, while the second stage targeted the wireless links and the transmitted power consumption problem. But such a study failed at considering the power consumption of the BBU pool and the cloud physical resources.

1.4 Chapter Summary

This chapter presents and explains the most relevant work related to C-RAN. It shows that the existing work did not consider the RRH selection for the UEs and the load balancing between BBUs and RRHs jointly. However, the effect of wrong RRH-Sector selection could lead to the loss of operator revenues and lower the QoS of the UEs. Moreover, the inappropriate mapping between the BBUs and RRHs incurs load imbalances in the C-RAN, which could lead to the

degradation of the NP. Therefore, in this chapter we introduced a framework that considers the RRH-Sector selection for the UEs as well as load balancing between BBUs and RRHs. This framework can better utilize the C-RAN resources from the operator and UE perspectives, which is not considered before up to our knowledge. Furthermore, it integrates the concept of SON and C-RAN towards a more centralized managed network. The next chapter presents the considered C-RAN system model with all the parameters needed to operate the network.

CHAPTER 2

C-RAN SYSTEM MODEL

2.1 Introduction

This chapter presents the considered C-RAN network model, its parameters and assumptions. First, it shows how the C-RAN topology is formed and what practical elements are used in the presented model, such as the technology used, etc. Moreover, it illustrates how users are distributed and which parameters are considered during the selection and load balancing steps. Finally, the most critical system constraints are determined.

2.2 System model

C-RAN architecture consists of three main components named BBUs, RRHs, and the front haul links. Figure 3 shows that each BBU serves a cluster of adjacent RRHs which represented by a specific sector, where each sector is defined by a different color that has its own set of resource blocks (RBs). The RB takes 180 kHz bandwidth in the frequency domain and 6 or 7 orthogonal frequency-division multiplexing (OFDM) symbols (i.e., one slot) in the time domain. The frequency section is comprised of 12 consecutive sub-carriers of 15 kHz. The front haul link between the BBUs and RRHs relies on switch fabric and optic links to form what is known as an optical transport network. The switch fabric is a network that consists of switches, multiplexers and optical splitters, while the optic links should be with low latency and high bandwidth. The SON server/controller realizes the self-organizing allocation of RRHs to sectors.

The considered C-RAN model consists of N RRHs that are assigned to K sectors. Each sector has total bandwidth $\hat{B}$. We assume that $\hat{B}$ is equals to 100 RBs since the C-RAN is operating with 20 MHZ. The sectors are served by B BBUs where each BBU can serve one or more sectors. RRH_i serves a set UM_i of M_i connected users, where $i = 1, \ldots, N$. The users are

divided into J classes differentiated by set of $\hat{n}$ criteria (e.g. data rate, delay, etc.), $Z = \{ z_1, z_2, \dots, z_{\hat{n}}\}$, with corresponding weights $W = \{ w_1, w_2, \dots, w_{\hat{n}}\}$ that are considered in RRH-Sector selection. Also, SoS_b denotes the set of sectors served by BBU_b. Equally, SoR_k represents the set of RRHs constituting sector k. Assignment of RRHs to sector *k* at time period *t* is expressed as $R_k^t = \{R_{1k}^t, R_{2k}^t, \dots \dots, R_{Nk}^t\}, k$ where $R_{nk}^t = 1$ if RRH_n has been assigned to sector *k* at time period *t*. New assignment for each RRH at time period *t+1* is denoted as R_k^{t+1}. We assume that the considered C-RAN is owned by one operator. Moreover, we assume that the RRHs and Sectors collect statistics of the connection parameters such as the number of resource blocks (RBs) assigned for each UE that depends on the scheduler used, connection arrival rates, connection duration, and the respective QoS metrics (data rate, delay, etc.). The general model considers that the connections can demand different number of RBs. Nevertheless, in this thesis, to simplify the presentation, we assume in the following that each connection associated to a specific sector requires the same number of RBs, but the presented methodology is not limited to this case as explained in CHAPTER 3. As shown in Figure 2.1, we model each sector with multiple servers which can serve *m* maximum number of UEs simultaneously with maximum amount of used bandwidth $\hat{B}$ (i.e., RBs). Class j UEs connection requests are characterized by requested QoS metrics (data rate and delay, etc.), connection arrival rate λ_j, service rates μ_j, and reward parameter r_j. A class *j* connection served in the RRH-Sector provides a reward r_j to the operator. This parameter can have monetary interpretation or can be treated as a control parameter used to establish the connection priority.

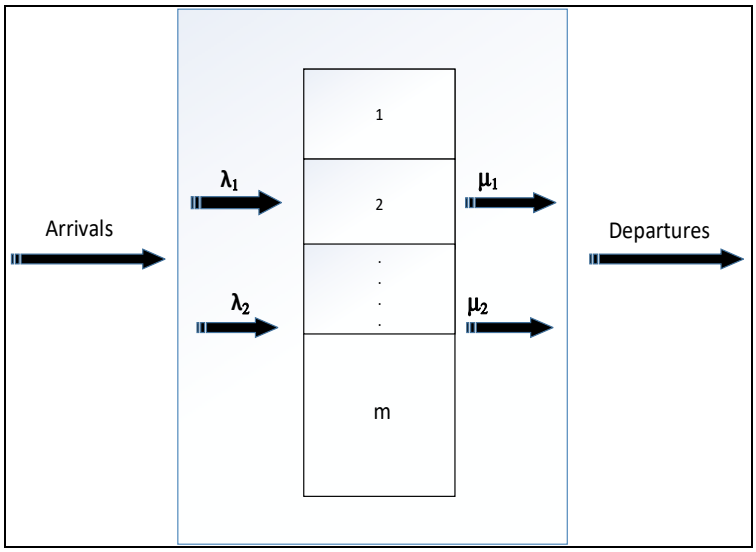

Figure 2.1 Sector with multiple servers

In LTE, each UE monitors two types of synchronization signals for sector search: primary synchronization signals (PSS) and secondary synchronization signals (SSS). First, each UE monitors the PSS to accurately acquire slot synchronization within 5 ms and identify the sector number. Second, each UE monitors the SSS to acquire frame synchronization and identify the sector group number (Sriharsha et al., 2017). Therefore, we assume that the RRH-Sector selection process takes place over a frame of 10 ms, as shown in Figure 2.2. Each BBU has a scheduler that allocates the resource blocks (RBs) to the UEs in each sector. In the considered C-RAN users can have different numbers of RBs depending on the type of the scheduler. There are different types of schedulers that could or could not consider the channel state information before the scheduling decisions. In this thesis, we use the Frequency Domain Round Robin (FD-RR) scheduler to schedule the resources (i.e.,$\hat{B}_{RBs}$) among the number of connections (i.e., UEs) in each sector. Since the RRH-Sector selection model is based on the MDP, which represents the state of each sector by the number of connections, we selected the FD-RR scheduler since it relates the number of connections to the number of RBs in the system in a simple way. However, the proposed approach could be used if a different type of scheduler is used that assigs different number of RBs to the users associated to a specific sector. This is presented in (Dziong et al., 1990) where an adaptive bandwidth allocation scheme is considered. Each sector has its own set of equal bandwidth blocks that are referred to as

resource blocks (RBs). The scheduler allocates the RBs to the UEs cyclically regardless of the traffic requirements and the channel conditions while handling all connections with the same priority. As a result, each connection that comes to a specific sector is going to obtain the same number of resources as all current connections (i.e., UEs) allocated to that specific sector. Since we assume that FD-RR uses Resource Allocation Type 0 and the system bandwidth is 20 MHZ, the resource blocks are divided into multiple groups. Each group is called the Resource Block Group (RBG) and consists of 4 RBs (*LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures*). Note that the maximum number of users that could be scheduled simultaneously depends on the number of the Physical Downlink Control Channels (PDCCHs) that carry downlink control and scheduling assignments information (Holma et al., 2009; Love et al., 2008; *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures*). For simplicity, we did not consider the number of PDCCHs in this thesis because it is a parameter configured by the network operator. Therefore, each sector could schedule up to 250 concurrent UEs during a given frame that is calculated by multiplying the 100 RBs by 10 subframes divided by the RBG size, which is 4. Hence, this indicates that each UE is going to get a minimum of 4 RBs when the sector reaches its maximum capacity.

In our approach, the RRH-Sector selection for a new connection depends on the MDP model. The selection process is completed for all UEs during each frame (i.e., 20 time slots, each of them is 0.5 ms). Then, the focus is to determine the suitable mapping for logical connections between BBUs, RRHs, and Sectors. The server/controller balances the load and improve the NP at time period t+1 under a given network condition represented by connected UEs at time period t based on selected KPIs. The server/controller is responsible for identifying the KPIs' status by compiling the desired information from the UEs. Each UE performs Channel Quality Indicator (CQI), Pre-coding Matrix Indicator (PMI), and Rank Indicator (RI) measurements and reports them on a periodic or aperiodic basis. Detailed information about the CQI report is discussed in (*LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures*). Here, we assume that a periodic CQI reporting is used. The CQI reporting can be configured according to the operator preference (e.g., every 40 subframes which is 40 ms) as shown in Figure 2.2. Therefore, the load balancing algorithm's execution by the SON

server/controller takes place in each time period t, which has to be bigger than the CQI reporting.

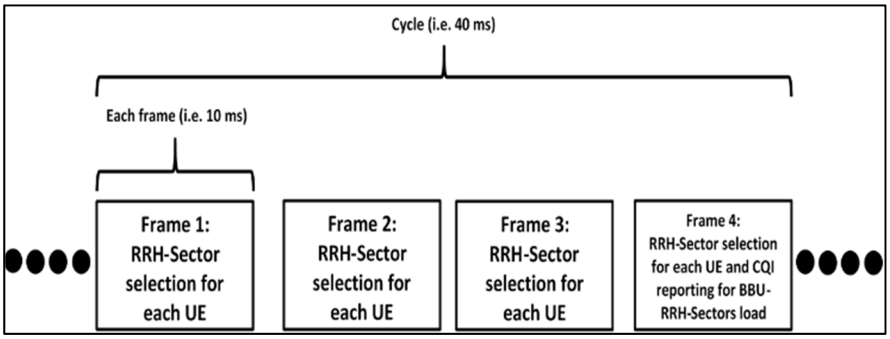

Figure 2.2 Cycle and frame structure

The proposed framework optimizes the RRH-Sector selection and network load balancing based on the centralized SON controller. Such an architecture introduces some challenges related to possible failures of the controller, time constraints for changing the RRH-Sector-BBU mapping, and constraints on the processing delay of sub-frame between the connection of BBUs and RRHs (Checko et al., 2015). Possible solutions for these challenges are presented in (Arslan et al., 2015; Tsagkaris et al., 2015) .

2.3 C-RAN constraints

This thesis introduces a C-RAN architecture with a centralized-SON feature. The proposed framework provides optimization of the RRH-Sector selection and network load balancing. However, there are still some constraints related to the model that must be considered:

- Since the server or controller acts as a focal point and is responsible for managing of the BBU pool, the network may break down if the Host manager malfunctions.
- The switching components used in configuring the RRH-Sector-BBU mapping should not influence the time scale of the sub-frames (i.e., 1ms).
- The processing delay of the sub-frame between the connection of BBUs and RRHs should be less than 1 ms, to satisfy the Hybrid automatic repeat request (HARQ) constraint (Checko et al., 2015).

Authors in (Arslan et al., 2015; Tsagkaris et al., 2015) present possible solutions for the introduced limitations and challenges above.

2.4 Chapter Summary

This chapter introduced the C-RAN system model when LTE is applied. We assume that in the area served by the C-RAN the RRHs are clustered into several sectors. The sectors are further handled by several BBUs, where each BBU can serve one or more sectors. Each sector serves a set of connected users that are further differentiated according to their class. All the necessary parameters and assumptions to operate the C-RAN are presented and discussed. The next chapter introduces the first part of the framework that addresses the RRH-Sector selection for the UEs.

CHAPTER 3

RRH-SECTOR SELECTION MODEL BASED ON MDP WITH PENALTY

3.1 Introduction

In this chapter, we propose an RRH-Sector selection model based on operator and user utilities. This model also can be considered as a short-term load balancing based on the current state of the sectors since the users located at the boundary of two sectors can be allocated to either of them. The operator utility is the average reward from the network and its maximization is based on sector shadow price and sector net-gain concepts derived from the MDP decomposition approach presented in (Dziong et al., 1990). Then, we introduce the user utility defined as the users' satisfaction from obtained QoS metrics. The user utility is calculated using the sigmoid function. Then, we introduce the operator reward penalty that is a function of the user utility. By adding this penalty to the MDP based model that maximizes the operator utility, we integrate the operator and user utilities in the MDP-P model.

3.2 Operator utility calculation

The operator objective is to find the optimal RRH-Sector selection policy π that maximizes the operator utility defined as the average reward from the network:

$$\acute{R}(\pi) = \sum_{j \in J} \bar{\lambda}_j r_j = \sum_{k \in K} \sum_{j \in J} \bar{\lambda}_j^k \, r_j \tag{3.1}$$

where $\bar{\lambda}_j$ is the rate of class j accepted connections in the network, r_j is the average operator reward from carrying a class j connection, and $\bar{\lambda}_j^k$ is the rate of class j accepted connections in sector k.

3.2.1 Sector shadow price concept and net gain

The sector shadow price can be defined as the expected loss of future reward from connections rejected due to the acceptance of a new connection demand in a given Sector in the current network state. Let $p_j^k(z,\pi)$ denote the shadow price in our system for a class j connection accepted in Sector k in network state $z = [z_j^k]$ under sector selection policy π, where z_j^k denotes the number of class j connections in sector $k \in K$. Then we can define the net-gain from accepting a class j connection in sector k in state z as:

$$g_j^k(z,\pi) = r_j - p_j^k(z,\pi) \tag{3.2}$$

The advantage of expressing the net-gain as the function of shadow price is that the shadow price values are independent of the new connection's reward, and therefore the policy storage memory can be reduced if there are several connection classes differentiated only by the reward parameter. Note that by assuming a Poisson arrival process, the system can be modeled as a Markov Decision Process. Then, it can be shown that the optimal RRH-Sector selection policy, which maximizes the average reward from the network, can be obtained by the following algorithm corresponding to the policy iteration algorithm (Dziong et al., 1990)

1. In the network operating under the given RRH-Sector selection policy π, defined by $p_j^k(z,\pi)$, estimate the Markov process parameters and, using these values, compute the improved values of sector shadow prices $p_j^k(z,\pi')$.

2. For each connection demand of class j, implement the improved policy by selecting the RRH-Sector pair that offers the maximum sector net-gain over all possible sectors:

$$g_{max} = \max_{k\in K} g_j^k(z,\pi') = \max_{k\in K}\left[r_j - p_j^k(z,\pi')\right] \tag{3.3}$$

   If the net-gain is negative, reject the demand.
   Go back to step 1.

Note that the number of network states z is a product of numbers of possible states in each sector, so the number of states z can be very large. Therefore, the calculation of shadow prices $p_j^k(z,\pi)$ from the exact MDP model is not practical for a realistic network. To cope with this issue, we propose to decompose the network Markov process into a set of independent Sector Markov processes by assuming that connection demands form an independent Poisson process in each sector. In this case, the Markov process in sector k is defined by the state-dependent arrival rate $\lambda_j^k(x,\pi)$ and departure rate μ_j, where $x = [x_j]$ denotes the sector state and x_j is the number of class j users connected to the sector. This approach is analogous to the framework presented in (Dziong et al., 1990) for wired mesh networks, where it is shown that for realistic networks, the independence assumption error is negligible. After decomposition of the Markov process, the sector shadow prices $p_j^k(x,\pi)$ can be calculated independently for each sector, and the policy iteration algorithm is modified as follows:

1. In the network operating under given sector selection policy π, defined by $p_j^k(x,\pi)$, estimate the Markov process parameters, $\lambda_j^k(x,\pi)$ and μ_j, and using these values compute the improved values of sector shadow prices $p_j^k(x,\pi')$.

2. For each class j connection demand, implement the improved policy by selecting the RRH-Sector pair that offers the maximum sector net-gain over all possible sectors:

$$g_{max} = \max_{k\in K} g_j^k(x,\pi') = \max_{k\in K}\left[r_j - p_j^k(x,\pi')\right] \tag{3.4}$$

   If the net-gain is negative, reject the demand.
   Go back to step 1.

3.2.2 Sector shadow price calculation in the decomposed model

In the decomposed model, for a given RRH-Sector selection policy π, sector k reward process can be described independently by the set $\{r_j, \lambda_j^k(x,\pi), \mu_j\}$ so that one can define sector k net-gain $g_j^k(x,\pi)$ as the expected reward from accepting class j connection at sector k in state x.

The reward rate from all connections in the sector at state x is given by:

$$q(x) = \sum_{j \in J} r_j x_j \mu_j \tag{3.5}$$

Then, for given values of $\lambda_j^k(x,\pi)$ and μ_j, the net-gains $g_j^k(x,\pi)$ and corresponding shadow prices $p_j^k(x,\pi)$ can be obtained by applying the value iteration algorithm (Schweitzer et al., 1979). Since the value iteration algorithm is developed for discrete-time Markov processes, its application to our continuous-time Markov process requires a uniformization of the state sojourn times to an average time τ (Grassmann, 1977). After the uniformization, the value functions $V_i^k(x,\pi)$ for our system can be computed from the following recurrence relation:

$$\begin{aligned} V_i^k(x,\pi) = q(x)\tau &+ \sum_{j \in J^k} \lambda_j^k(x,\pi)\tau \left[V_{i-1}^k\left(x+\delta_j,\pi\right) - V_{i-1}^k(x,\pi)\right] \\ &+ \sum_{j \in J} x_j \mu_j \tau \left[V_{i-1}^k\left(x-\delta_j,\pi\right) - V_{i-1}^k(x,\pi)\right] \\ &+ V_{i-1}^k(x,\pi), k \in K \end{aligned} \tag{3.6}$$

where i is the iteration index and δ_j is a J-dimension vector with 1 at j position and 0 in all other positions, $+\delta_j$ represents the arrival of class j user at the considered sector and $-\delta_j$ represents the departure of class j users from the considered sector. Once the value iteration algorithm converges, the sector net-gain values are given by:

$$g_j^k(x,\pi) = \lim_{i \to \infty} \left[V_i^k\left(x+\delta_j,\pi\right) - V_i^k(x,\pi)\right] \tag{3.7}$$

and the corresponding sector shadow price values are given by:

$$p_j^k(x,\pi) = r_j - g_j^k(x,\pi) \tag{3.8}$$

As mentioned before, the net-gain values are used to maximize the operator utility in the RRH-Sector selection model. Note that, as mentioned before in this paper the shadow price calculation assumes that that the connections associated with a specific sector require the same amount of bandwidth (RBs). However, the proposed methodology is valid also for the cases where the connections demand different number of RBs. In this case, the calculation of the shadow prices for connections with variable number of RBs can be done by using models similar to the ones presented in (Dziong et al., 1990).

3.3 User utility calculation

For each RRH-Sector pair candidate, the user utilities for each criterion is evaluated. The utility for a given criterion is a normalized numeric value in the range of [0, 1] which indicates the user's satisfaction with respect to the criterion value provided by the RRH-Sector pair.

In the RRH-Sector selection context, to compute the user utilities, we use the sigmoid function presented in (Nguyen-Vuong et al., 2008) and illustrated in Figure 3.1. The sigmoid function has several good features such as twice differentiability, increasing function, concavity, and convexity conditions (Nguyen-Vuong et al., 2008).

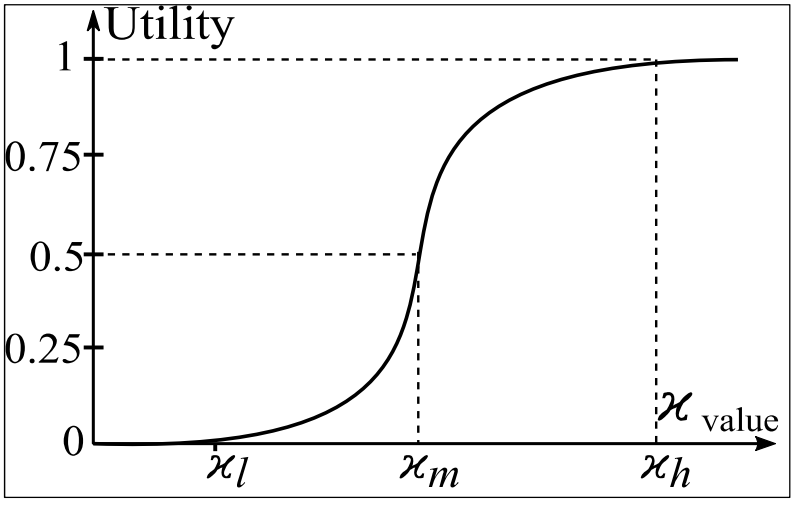

Figure 3.1 Sigmoid function for user utility calculation

The following equation, corresponding to the sigmoid function, expresses the utility value calculation for a given criterion χ:

$$u(\chi) = \begin{cases} 0, \chi_n < \chi_l \\ \dfrac{\left(\dfrac{\chi_n - \chi_l}{\chi_m - \chi_l}\right)^{\zeta}}{1 + \left(\dfrac{\chi_n - \chi_l}{\chi_m - \chi_l}\right)^{\zeta}}, \chi_l \leq \chi_n \leq \chi_m \\ 1 - \dfrac{\left(\dfrac{\chi_h - \chi_n}{\chi_h - \chi_m}\right)^{\gamma}}{1 + \left(\dfrac{\chi_h - \chi_n}{\chi_h - \chi_m}\right)^{\gamma}}, \chi_m \leq \chi_n \leq \chi_h \\ 1, \chi_n > \chi_h \end{cases} \quad (3.9)$$

where the parameters are defined as follows:

$\zeta \geq max\left\{\frac{2(\chi_m - \chi_l)}{\chi_h - \chi_m}, 2\right\}$: tuning coefficient for the shape steepness parameter,

$\gamma = \frac{\zeta(\chi_h - \chi_m)}{\chi_m - \chi_l}$: shape steepness parameter,

χ_n : value obtained for criterion χ,

χ_l : minimum acceptable value for criterion χ,

χ_h: maximum desired value for criterion χ,

χ_m: satisfaction and non-satisfaction frontier.

Equation (3.9) is for an upward criterion. For downward criteria, we use $1 - u(\chi)$. An upward criterion gives greater utility for greater metric value (e.g., data rate) while a downward criterion gives greater utility for lower metric value (e.g., delay). To obtain the aggregated class j user utility connected to sector k through RRH n, for all the considered criteria we use the multiplicative aggregation expressed by:

$$U_j^{k,n}(C, W) = \prod_{i=1}^{\hat{n}} \left[u_j^{k,n}(c_i)\right]^{w_i}, j \in J, k \in K, n \in N, \sum_i w_i = 1 \quad (3.10)$$

The multiplicative aggregation is well suited for our multi-criteria decision because, as noted in (Nguyen-Vuong et al., 2008), it takes care of the interdependence of the criteria. Moreover, it conserves against ignoring completely specific criteria with lower weights.

3.4 Criteria considered for RRH-sector selection in the user utility

3.4.1 RSS

We use the free space propagation model, a derivative of the Friis free space equation (Shaw, 2013), to compute the RSS. The RSS depends on the distance between UE and the RRH and the user transmission powers. Then the RSS is expressed by:

$$P_{R(\varepsilon,n,k)} = \frac{P_t \times G_t \times G_r \times \gamma^2}{(4 \times \pi)^2 \times d^{\alpha} \times \omega} \tag{3.11}$$

where

$P_{R(\varepsilon,n,k)}$, P_t : received and transmitted powers,

G_t, G_r: transmitter and receiver antenna gains,

γ: signal wavelength,

d: the distance between transmitter and receiver,

α, ω: path loss coefficient and system loss factor.

3.4.2 Data rate

The data rate offered by sector k through RRH n depends on the user location (i.e., signal-to-interference-plus-noise ratio (SINR)). Therefore, it is very important to select the best RRH-Sector pair that can satisfy the user requirement and efficiently utilize the resources in the network at the same time.

The SINR can be represented as:

$$SINR_{\varepsilon,n,k} = 10\ x \log \frac{P_{R(\varepsilon,n,k)}}{\widetilde{N} + \tilde{I}} \tag{3.12}$$

where $P_{R(\varepsilon,n,k)}$ is the RSS for user ε connected to sector k through RRH n, $\tilde{I}$ is the interfering power comes from all RRHs not belonging to sector k, and $\widetilde{N}$ is the average noise power.
It is assumed that the best modulation coding scheme (MCS) could be adopted for a given SINR, that offers the highest data rate. Therefore, the offered data rate for user ε connected to sector k through RRH n can be represented by the Shannon formula:

$$\hat{d}_{\varepsilon,n,k} = \hat{B}_{\varepsilon,n,k}.\, log_2(1 + SINR_{\varepsilon,n,k}) \tag{3.13}$$

where $\hat{B}_{\varepsilon,n,k}$ is the total bandwidth of allocated RBGs to UE ε connected to sector k through RRH n from the FD-RR scheduler during a specific frame (i.e., $\hat{B}_{\varepsilon,n,k} = \sum RBG_{\varepsilon,n,k}$).

3.4.3 Service cost for the user

The cost of the service is calculated according to the total bandwidth of allocated RBGs during a specific frame by the FD-RR scheduler at a given sector and is expressed by:

$$C_B = \hat{B}_{\varepsilon,n,k}\ x\ c_{RBG} \tag{3.14}$$

where c_{RBG} is the price per one RBG.

3.5 Operator reward penalty calculation

In the hardware implementation, the computation of a sigmoid function is considered as one of the constraint factors. Therefore, for reducing the complexity in the implementation, the Piecewise Linear Approximation function which represents the operator reward penalty is widely being used for realizing the sigmoid function into hardware (Ngah et al., 2017). The operator reward penalty rate, $c_j^{k,n}\left(U_j^{k,n}\right)$ is introduced here which is a function of the user utility U_j^k defined in subsection 3.2. To design the $c_j^{k,n}\left(U_j^{k,n}\right)$ function, we divide the user

utility values into zones, as suggested in (Sevcik, 2002). These are the frustration zone, tolerance zone, and satisfaction zone, as illustrated in Figure 3.2. The thresholds delimiting the zones denoted as $\mathcal{T}_1$ and $\mathcal{T}_2$, can be set according to the parameter for which the utility is considered. In the frustration zone, we consider that the penalty decreases slowly, as the user utility increases, to express that the user remains unsatisfied below a certain threshold, here $\mathcal{T}_1$. In the tolerance zone, the slope of penalty decrease is high to express the user's willingness to pay for service as he now has the minimum requirement. In the satisfaction zone, the penalty decrease has a small slope again because the user's satisfaction does not increase much after reached a certain satisfaction level delimited by a threshold $\mathcal{T}_2$. We also introduce constants $\hat{k}$, $\hat{k}_1$ and $\hat{k}_2$, in range [0, 1], that allow us to express the penalty as a reasonable fraction of the reward parameter. $\hat{k}$ ensures that the operator receives an amount of reward even if user utility is equal to zero since the user has used the network. $\hat{k}_1$ and $\hat{k}_2$ delimit the proportion of $r_j\mu_j$ for the tolerance and satisfaction zones. According to the above remarks, $c_j^{k,n}\left(U_j^{k,n}\right)$ is represented by a piecewise function shown in Figure 3.2 that is defined as follows:

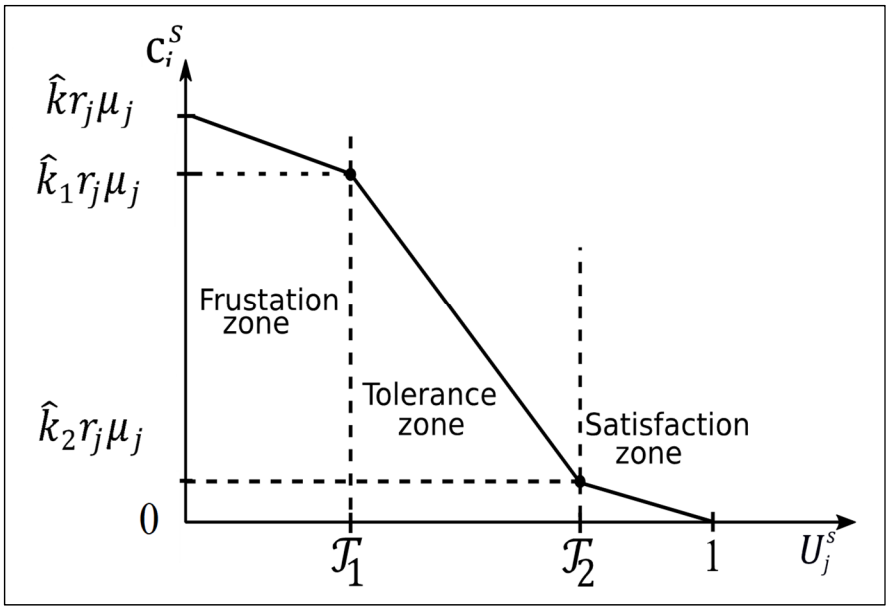

Figure 3.2 Variation of penalty as a function of user utility

$$c_j^{k,n}\left(U_j^{k,n}\right) = \begin{cases} \dfrac{r_j\mu_j\left(\hat{k}_1 - \hat{k}\right)}{\mathcal{T}_1} U_j^{k,n} + \hat{k} r_j \mu_j , & U_j^{k,n} \in [0, \mathcal{T}_1[\\ r_j\mu_j \left[\dfrac{\left(\hat{k}_1 - \hat{k}_2\right)\left(U_j^{k,n} - \mathcal{T}_1\right)}{\mathcal{T}_1 - \mathcal{T}_2} + \hat{k}_1 \right], & U_j^{k,n} \in [\mathcal{T}_1, \mathcal{T}_2[\\ \dfrac{\hat{k}_2 r_j \mu_j}{\mathcal{T}_2 - 1}\left(U_j^{k,n} - 1\right), & U_j^{k,n} \in [\mathcal{T}_2, 1[\end{cases} \tag{3.15}$$

3.6 Policy iteration with added user utility penalty

The integration of both user and operator utilities within MDP is done by reducing the reward rate from the connection by a penalty rate that is a function of the user utility value related to the selected RRH-Sector pair. Therefore, the higher the value of the user's utility, the smaller the penalty will be. In this case, the class j connection reward rate for connections accepted at the RRH located in sector k is defined as:

$$q_j = r_j\mu_j - c_j^{k,n}\left(U_j^{k,n}\right) \tag{3.16}$$

where $c_j^{k,n}\left(U_j^{k,n}\right)$ is the user reward penalty rate being a function of the class j user utility for sector k.

The RRH-Sector selection objective is to find the optimal policy π that maximizes the mean value of reward from the network defined as:

$$\bar{R}(\pi) = \sum_{j \in J} \left(r_j - \overline{c_j}/\mu_j\right)\bar{\lambda}_j \tag{3.17}$$

where $\overline{c_j}$ and $\bar{\lambda}_j$ are the average reward penalty rate and the average arrival rate for accepted class j connections, respectively. Note that the mean value of reward from the network can be presented as a sum of average rewards from the sectors.

$$\bar{R}(\pi) = \sum_{k \in K} \sum_{j \in J} \left(r_j - \bar{c_j}^k / \mu_j\right) \bar{\lambda}_j^k \tag{3.18}$$

where $\bar{c_j}^k$ and $\bar{\lambda}_j^k$ are the average reward penalty rate and the average arrival rate for accepted class j connections in sector k, respectively.

Then the sector shadow price can be defined as the expected loss of future reward from connections rejected due to the acceptance of a new connection demand in a given sector in the current network state. Let $p_j^k(z, \pi)$ represents the shadow price for class j connection accepted in sector k in network state $z = \left[z_j^k\right]$ under RRH-Sector selection policy π. Then we can define the RRH-Sector pair net-gain from accepting a class j connection at RRH located in sector k in state z as follows:

$$g_j^{k,n}(z, \pi) = r_j - c_j^{k,n}\left(U_j^{k,n}\right)/\mu_j - p_j^k(z, \pi) \tag{3.19}$$

It is important to underline that while $c_j^{k,n}\left(U_j^{k,n}\right)$ depends on the new connection's location (RSS in the user's utility), the value of $p_j^k(z, \pi)$ does not since this is an expectation related to future connections. Then, assuming a Poisson arrival process, the system can be modeled as a Markov Decision Process and the optimal RRH-Sector selection policy, which maximizes the mean value of reward from the network defined by equation (18), is obtained from the following policy iteration algorithm:

1. In the network operating under the given RRH-Sector selection policy π, defined by $p_j^k(z, \pi)$, estimate the Markov process parameters and the expected average reward penalty rates (being a function of z). Then using these values compute the improved values of sector shadow prices $p_j^k(z, \pi^*)$.

2. For each class j connection demand, implement the improved policy by selecting the RRH-Sector pair that offers the maximum RRH-Sector net-gain over all possible sectors.

$$g_{max} = \max_{k,n \in K,N} \left[r_j - c_j^{k,n}\left(U_j^{k,n}\right)/\mu_j - p_j^k(z, \pi^*)\right] \tag{3.20}$$

If the net-gain is negative, reject the demand.

Go back to step 1.

3.6.1.1 Sector shadow price calculation in the decomposed model with added user utility penalty

Figure 3.3 illustrates the state transition diagram in the decomposed model for a sector that can admit five users, which are divided into two classes.

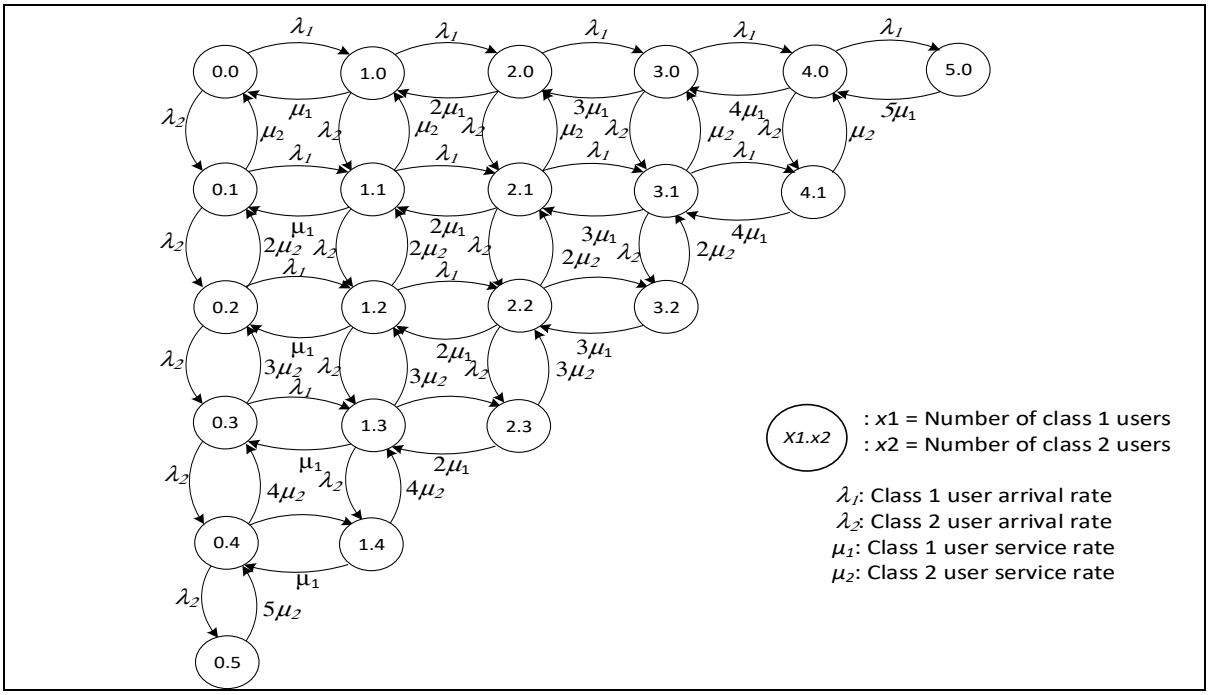

Figure 3.3 State transition diagram for a sector

$x \rightarrow x + \delta_j$: admission of class j user

$x \rightarrow x - \delta_j$: the departure of class j user

Note that the number of network states z is a product of numbers of possible states in each sector so the number of states z can be very large. Therefore, the calculation of shadow prices $p_j^k(z, \pi)$ from the exact MDP model is not practical for any realistic network. To cope with this issue, we propose to decompose the network Markov process into a set of independent

sector Markov processes by assuming that connection demands form an independent Poisson process in each sector. In this case, the Markov process of sector k is defined by the state-dependent arrival rate $\lambda_j^k(x,\pi)$, the departure rate μ_j, and the sector state denoted by $x = [x_j]$, where x_j is the number of class j users connected to the sector. This approach is analogous to the framework presented in (Dziong et al., 1990) for wired mesh networks, where it is shown that for realistic networks, the independence assumption error is negligible. Figure 3.3 illustrates the state transition diagram in the decomposed model for a sector that can admit 5 users, which are divided into 2 classes. In the decomposed model, for a given RRH-Sector selection policy π, sector k reward process can be described independently by the set $\{r_j, \lambda_j^k(x,\pi), \mu_j, \bar{c}_j^{\,k}\}$ so that one can define the sector k net-gain $g_j^k(x,\pi)$ as the expected reward from accepting class j connection at sector k in state x. The rate of reward from all connections in the sector when in state x is now given by:

$$q(x) = \sum_{j \in J} \left(r_j \mu_j - \bar{c}_j^{\,k}\right) x_j \tag{3.21}$$

Then, for given values of $\lambda_j^k(x,\pi)$ and μ_j, the sector net-gains $g_j^k(x,\pi)$ and the corresponding sector shadow prices $p_j^k(x,\pi)$ can be obtained by applying the value iteration algorithm (Schweitzer & Federgruen, 1979). Since the value iteration algorithm is developed for discrete-time Markov processes, its application to our continuous-time Markov process requires a uniformization of the state sojourn times to an average time τ (Grassmann, 1977). After the uniformization, the value functions $V_n^k(x,\pi)$ for our system can be computed from the following recurrence relation:

$$\begin{aligned} V_i^k(x,\pi) = q(x)\tau &+ \sum_{j \in J^k} \lambda_j^k(x,\pi)\tau \left[V_{i-1}^k\left(x+\delta_j,\pi\right) - V_{i-1}^k(x,\pi)\right] \\ &+ \sum_{j \in J} x_j \mu_j \tau \left[V_{i-1}^k\left(x-\delta_j,\pi\right) - V_{i-1}^k(x,\pi)\right] \\ &+ V_{i-1}^k(x,\pi), k \in K \end{aligned} \tag{3.22}$$

where i is the iteration index and δ_j is a J-dimension vector with 1 at j position and 0 in all other positions, $+\delta_j$ represents the arrival of class j user at the considered sector and $-\delta_j$ represents the departure of class j users from the considered sector.

Finally, the sector shadow prices $p_j^k(x,\pi)$ can be calculated independently for each sector, and the policy iteration algorithm is modified as follows:

1. In the network operating under the given RRH-Sector selection policy π, defined by $p_j^k(x,\pi)$, estimate the Markov process parameters, $\lambda_j^k(x,\pi)$ and μ_j, and the average reward penalty rates for accepted class j connections, $\bar{c}_j^k$. Then using these values compute the improved values of sector shadow prices $p_j^k(x,\pi^*)$.

2. For each class j connection demand, implement the improved policy by selecting the RRH-Sector pair that offers the maximum RRH-Sector net-gain over all possible RRH-Sector pairs set $\Re$.

$$g_{max} = \max_{(k,n)\in\Re}\left[r_j - c_j^{k,n}\left(U_j^{k,n}\right)/\mu_j - p_j^k(x,\pi^*)\right] \tag{3.23}$$

   If the net-gain is negative, reject the demand.

   Go back to step 1.

3.7 Chapter Summary

This chapter presents a model for the RRH-sector selection for a user demanding a new connection. The proposed RRH-sector selection model is based on the user utility as well as the operator utility, which represents the users' QoS and network profit, respectively. The proposed model uses the concepts of RRH-Sector shadow prices and RRH-Sector net gains derived from MDP decomposition with the objective of maximizing the operator-user utility. Then, we introduce the operator reward penalty that is a function of the user utility. By adding this penalty to the MDP based model that maximizes the operator utility, we integrate the operator and user utilities in the MDP-P model. In the next chapter, the second part of the framework dealing with RRH-Sector-BBU load balancing is addressed.

CHAPTER 4

DYNAMIC RRH- SECTOR-BBU MAPPING AND RELATED KPIs

4.1 Introduction

As mentioned in the previous chapter, the MDP model can be also considered as a short-term load balancing based on the current state of the sectors since the users at the boundary of two sectors can be allocated to either of them. However, it does not address long term imbalances in the sector loads. To address this issue, in this chapter, we introduce RRH-Sector-BBU mapping for load balancing approach based on the network performance (NP) optimization where the considered NP term is a function of several KPIs. Specifically, the focus is to determine the suitable mapping for logical connections between RRHs, Sectors and BBUs, that balances the sector loads and improves the NP at time period t+1 under a given network condition at time period t based on selected KPIs. The considered KPIs are a function of the standard deviation of the sector loads, number of forced handover blocked users, number of handovers, and the power consumption. These KPIs are used by the SON server to optimize the RRH-Sector-BBU mapping.

4.2 Load standard deviation KPI

The load offered to a sector k is defined as:

$$A_k = \sum_{j \in J} \frac{\lambda_j}{\mu_j} \tag{4.1}$$

where λ_j is the connection demand arrival rate of class j and μ_j is the service rate of class j, where $j \in J$. Note that all the sectors have the same maximum capacity.

We adopt this KPI for the standard deviation minimization, ψ, of the load level among the C-RAN sectors. It is calculated by using all active sectors' load distribution. The load standard deviation at time t+1 is expressed by:

$$\psi = \sqrt{\frac{\sum_{k=1}^{K}(A_k^{t+1} - \overline{A_k}^{t+1})^2}{K^{t+1} - 1}} \tag{4.2}$$

where A_k^{t+1} is the load of sector k and $\overline{A_k}$ is the average sector load over K^{t+1}active sectors. Then, one of the objectives of our approach is to minimize ψ to reach a well-balanced C-RAN network, and the KPI for standard deviation ψ is defined by:

$$KPI_{\psi} = \begin{cases} 1 & if\ \psi = 0 \\ \dfrac{1}{1+\psi} & otherwise \end{cases} \tag{4.3}$$

4.3 Handovers KPIs

To formulate the KPIs for handovers, let ρ_{ij} represent the handover probability of users from RRH_i to RRH_j, where $i, j \epsilon N$, and H_{ij} denotes the handovers from RRH_i to RRH_j defined as $H_{ij} = \rho_{ij} UM_i$. Several methods for real-time estimation of ρ_{ij} can be found in (Ge et al., 2016; Vu et al., 2014). In this thesis, we assume that the user distribution in every RRH coverage area is uniform, thus ρ_{ij} can be modeled to be inversely proportional to the distance between RRH_i and RRH_j i.e.,($\rho_{ij} = \frac{1}{D_{ij}}$). For example, as shown in Figure 4.1, the distance between the centers of two adjacent RRHs is assumed to be "d" (e.g., the distance between RRH 1 and RRH 2). Therefore, the distance between RRH 1 and RRH 9 (i.e., $x_{1,9}$) can be calculated as:

$$x_{1,9} = \sqrt{(2d)^2 + \left(\sqrt{3}d\right)^2} = \sqrt{7}d \tag{4.4}$$

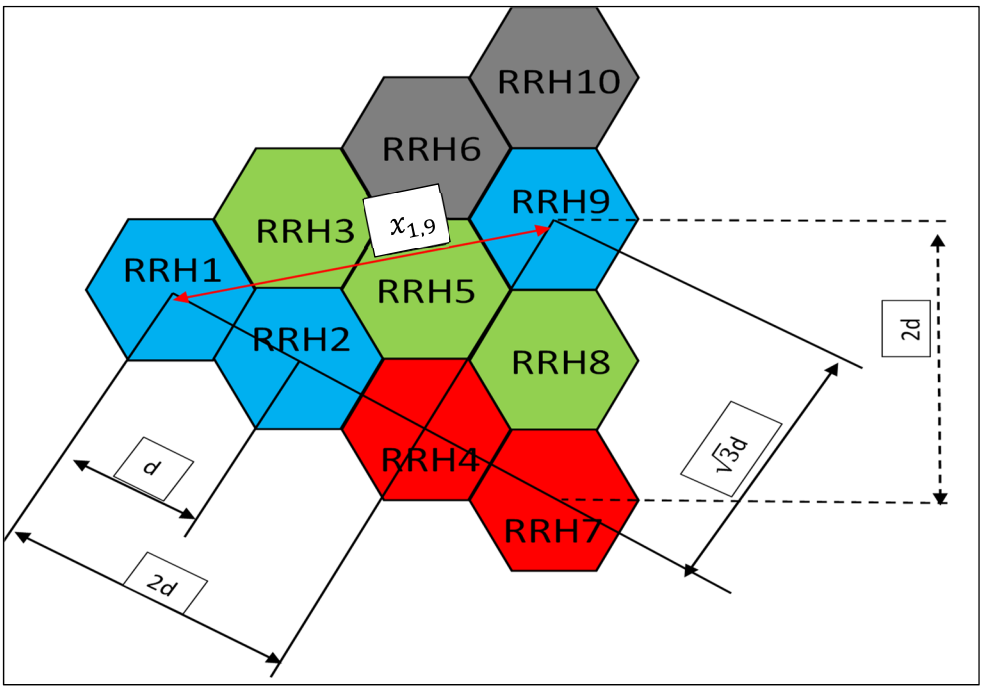

Figure 4.1 Distance calculation for RRH-Sector-BBU mapping

In the following we define three handover KPI types considered in this paper: inter-BBU handovers, intra-BBU handovers, forced handovers, and forced handover blocked users due to the mapping changes.

4.3.1 Inter-BBU handover KPI

The inter-BBU and intra-BBU handovers in C-RAN are corresponding to inter-eNodeB and intra-eNodeB handovers in LTE, respectively. Let Y_{ijb}^{t+1}t be a binary variable where $Y_{ijb}^{t+1} = 1$ in case RRH_i and RRH_j are linked with the same BBU_b at t+1 time period, i.e., $Y_{ib}^{t+1} = Y_{jb}^{t+1} = 1$, and $Y_{ib}^{t+1} = \sum_{k \in SOS_b} R_{ik}^{t+1}$. $Y_{ijb}^{t+1}$is used to evaluate the number of inter-BBU handovers where $Y_{ijb}^{t+1} = 1 - \sum_b Y_{ijb}^{t+1}$. $Y_{ij}^{t+1} = 1$ in case RRH_i and RRH_j are linked to different $BBUs$ at

t+1 time period. Then, the number of inter-BBU handovers at t+1 time period is represented by:

$$interBBU_{HO} = \sum_{i} \sum_{j \neq i} H_{ij} Y_{ij}^{t+1} \tag{4.5}$$

where $H_{ij} = \rho_{ij} UM_i$ indicates handovers from RRH_i to RRH_j and $\rho_{ij} = \frac{1}{D_{ij}}$, and since $Y_{ij}^{t+1} = f_1(Y_{ijb}^{t+1})$, therefore equation (4.6) can be rewritten as:

$$interBBU_{HO} = \sum_{i} \sum_{j \neq i} \rho_{ij} UM_i (1 - \sum_{b} Y_{ijb}^{t+1}) \tag{4.7}$$

Given that $Y_{ijb}^{t+1} = f_2(Y_{ib}^{t+1}. Y_{jb}^{t+1})$, where the dot (.) express the AND logical operation, thus:

$$interBBU_{HO} = \sum_{i} \sum_{j \neq i} \rho_{ij} UM_i (1 - \sum_{b} (Y_{ib}^{t+1}. Y_{jb}^{t+1})) \tag{4.8}$$

since $Y_{ib}^{t+1} = f_3(R_{ik}^{t+1})$, and $Y_{jb}^{t+1} = f_4(R_{jk}^{t+1})$, therefore, the term Y_{jb}^{t+1} is a function on the key binary term $R_{iK}^{t+1}, i = 1, \ldots, N$ i.e., $Y_{ij}^{t+1} = f_1\left(f_2\left(f_3\left(f_4(R_{ik}^{t+1})\right)\right)\right)$. Then the $\text{interBBU}_{\text{HO}}$ is defined as:

$$interBBU_{HO} = \sum_{i} \sum_{j \neq i} \frac{UM_i}{D_{ij}} (1 - (\sum_{b} \sum_{k \in SOS_b} (R_{ik}^{t+1}. R_{jk}^{t+1}))) \tag{4.9}$$

where the dot (.) expresses the AND logical operation. Finally, the interBBU handovers KPI (KPI_{inter}) is defined as:

$$KPI_{inter} = \begin{cases} 1 & if\ interBBU_{HO} = 0 \\ \dfrac{1}{1 + interBBU_{HO}} & otherwise \end{cases} \tag{4.10}$$

4.3.2 Intra-BBU handovers KPI

When users must change from one sector to another one within one BBU, intra-eNodeB handovers take place. Let Z_{ijk}^{t+1}be a binary variable such that $Z_{ijk}^{t+1} = 1$ in case sector k contains RRH_i and RRH_j at t+1 time period (i.e., $R_{ik}^{t+1} = R_{jk}^{t+1}$=1). The intra-BBU handovers is specified by using two different binary terms: $Z_{ij}^{t+1}$and Y_{ij}^{t+1}. $Z_{ij}^{t+1} = 1$ in case RRH_i and RRH_j are linked to two different sectors at t+1 time period and is defined as $Z_{ij}^{t+1} = 1 - \sum_k Z_{ijk}^{t+1}$. $Y_{ij}^{t+1} = 0$ if the same BBU is serving RRH_i and RRH_j. Thus, at t+1 time period the number of intra-BBU handovers ($intraBBU_{HO}$) is represented by:

$$intraBBU_{HO} = \sum_i \sum_{j \neq i} H_{ij}(Z_{ij}^{t+1} - Y_{ij}^{t+1}) \tag{4.11}$$

Since $Z_{ij}^{t+1} = f_5(Z_{ijk}^{t+1})$, thus:

$$intraBBU_{HO} = \sum_i \sum_{j \neq i} \rho_{ij} UM_i \left[(1 - \sum_s Z_{ijk}^{t+1}) - (1 - \sum_n Y_{ijb}^{t+1})\right] \tag{4.12}$$

Then knowing that $Z_{ijk}^{t+1} = f_6(R_{ik}^{t+1}, R_{jk}^{t+1})$, intra-BBU handovers can be rewritten as:

$$intraBBU_{HO} = \sum_i \sum_{j \neq i} \frac{UM_i}{D_{ij}} [(1 - \sum_k (R_{ik}^{t+1}.R_{jk}^{t+1})) - (1 - \sum_b \sum_{k \in SOS} (R_{ik}^{t+1}.R_{jk}^{t+1}))] \tag{4.13}$$

Since the binary term Z_{ij}^{t+1} depends on the principal binary term $R_{ik}^{t+1}, i = 1, 2, \dots, N$. i.e., $(Z_{ij}^{t+1} = f_5(f_6(R_{ik}^{t+1}, R_{jk}^{t+1})))$, then we have:

$$intraBBU_{HO} = \sum_i \sum_{j \neq i} \frac{UM_i}{D_{ij}} [\sum_b \sum_{k \in SOS} (R_{ik}^{t+1}.R_{jk}^{t+1}) - \sum_k (R_{ik}^{t+1}.R_{jk}^{t+1})] \tag{4.14}$$

Note that a critical constraint is that only one BBU should serve an RRH at *t+1* time period, i.e., $\sum_{b=1}^{B} R_{ib}^{t+1} = 1$. Consequently, the KPI of intraBBU handovers is given by:

$$KPI_{intra} = \begin{cases} 1 & if\ intraBBU_{HO} = 0 \\ \dfrac{1}{1 + intraBBU_{HO}} & otherwise \end{cases} \tag{4.15}$$

4.3.3 Forced handovers KPI

When a new RRH-Sector-BBU mapping is implemented to optimize the network performance, each RRH can change its sector mapping. When an RRH changes its sector mapping, its users need also to follow this change of sector association. Let V_{ik} be a binary variable with value $V_{ik} = 1$ in case RRH_i changed its existing sector to a new sector when the mapping is updated between time period t and time period t+1 (i.e., $V_{ik} = 1, if R_{ik}^{t} = 0\ and\ R_{ik}^{t+1} = 1$). Then, the forced handover number is given by:

$$f_{HO} = \sum_{k} \sum_{i} V_{ik}\, UM_i \tag{4.16}$$

where the binary term V_{ik} depends on the binary term R_{ik}^{t} and R_{ik}^{t+1} , $\forall i = 1,2, \ldots\ldots, N$, i.e., $(V_{ik} = f_7(R_{ik}^{t}, R_{ik}^{t+1}))$. Therefore the f_{HO} expression can be rewritten as:

$$f_{HO} = \sum_{k} \sum_{i} (R_{ik}^{t} + R_{ik}^{t+1}) UM_i \tag{4.17}$$

where the (+) operator is the OR logical operation. Thus, the KPI of the forced handover (KPI_f) is defined as:

$$KPI_f = \begin{cases} 1 & if\ f_{HO} = 0 \\ \dfrac{1}{1 + f_{HO}} & otherwise \end{cases} \tag{4.18}$$

4.3.4 Key Performance Indicator for blocked Users

Blocked users are the users that are deprived of network services because of the maximum capacity when Forced handover blocking occurs when the number of users connected to a specific sector exceeds the maximum capacity of that sector due to forced handovers resulting from the RRH-Sector-BBU mapping change. The number of forced handover blocked users at time period t+1 can be defined by:

$$BC = \sum_k max\left[\left(\left(\sum_i UM_i R_{ik}^{t+1}\right) - HC_k\right), 0\right] \tag{4.19}$$

where $i = 1, \ldots, N,\ \ k = 1, \ldots, K$, $R_{ik}^{t+1} = 1$ if RRH_i is assigned to sector k, UM_i represents the number of active users being handled by RRH_i and HC_k is the sector k maximum capacity. The $\mathrm{KPI_{BC}}$ is assumed to be 1 if there is no forced handover blocked users. Then the KPI for Forced handover blocking UEs can be expressed as:

$$KPI_{BC} = \begin{cases} 1 & if\ BC = 0 \\ \dfrac{1}{1 + BC} & otherwise \end{cases} \tag{4.20}$$

4.4 Key Performance Indicator for Power consumption

This subsection introduces the essential features required to determine C-RAN power consumption. There are two models to calculate the power model in the C-RAN. The first one

is the components power model while the second one is the Parameterized and linear power model (PM). The components power model is very complex because it has many details which might not have a big effect, but the Parameterized and linear power model is simple and applicable because it covers important aspects such as transmission bandwidth and the number of radio chains. More details regarding the involved factors to calculate the power model in C-RAN are explained in (Alhumaima et al., 2016).

4.4.1 Components PM

In general, the components power model in C-RAN is affected by three main parts and described below.

4.4.1.1 A power model for RRH

Each RRH in C-RAN consists of antenna arrays and RF transceivers where each one of them has its own power amplifier (P_{PA}). The power efficiency (η_{PA}) of the amplifier is the main element to affect the PA power consumption. Then the power consumed by the PA can be represented as:

$$P_{PA} = \frac{P_{TX}}{\eta_{PA}(\sigma_{feed})} \tag{4.21}$$

where P_{TX} is the PA output power that relies on the bandwidth share (i.e. the real transmitted number of symbols) and the antenna output power. σ_{feed} indicates the losses of the feeder. Also, each RRH has RF transceiver units that are responsible for modulation and demodulation of the signals, AC-DC and DC-ac conversions, and amplification of the gain. Then the RRH power consumption can be represented by:

$$P_{RRH} = \sum_{l=1}^{L} (P_{PA} + P_{RF}) \tag{4.22}$$

where *L* indicates the number of antenna/RF arrays, here we assume that each RRH is equipped with one antenna.

4.4.1.2 BBU power model.

Various functions are performed by the BBU which comprises of RBs Scheduling, forward error correction, filtering, fast Fourier transform (FFT) and OFDM specific processing, modulation and demodulation, and functions related to transport link, etc. These components can be measured in Giga operation per second (GPOS) and, after that, converted into power figures. The estimated power cost of a very large BBU is approximately 40 GOPS per watt (L. Liu & Yu, 2017). Then the BBU power can be represented as

$$P_{BBU} = \sum_{i \in I_{BB}} P_{i,BBU}^{ref} \, L^{x_i^L} \widehat{W}^{x_i^{\widehat{W}}} \tag{4.23}$$

where $P_{i,BBU}^{ref}$ in watts indicates the BBU power consumption with reference to the functions of the BBU. L is the number of antenna chains associated with RF transceivers with x_i^L scaling indicator. $\widehat{W}$ is the total bandwidth share used during the transmission with scaling indicator $x_i^{\widehat{W}}$. The author in (Cunhua et al., 2017) model BBU operations with exact scaling components and reference values to calculate BBU power consumption.

4.4.1.3 Optical transceiver power model.

The C-RAN network architecture has some challenges, and one of these is the front haul requires low latency with high bandwidth for transport networks. Many criteria affect the operation of the optical transceivers like the conditions of operations, the used technology, and the required output power, which consequently influence the consumed power.

Mainly the optical transceivers might be divided into two components. The first one is the optical transmitter component, where optical carriers are used to modulate the OFDM electrical signals by using direct or external modulated lasers. The second one is the receiver component that discovers the optical OFDM signals, whether through coherent or direct detection. Then the optical transceiver power consumption as introduced in (Wang et al., 2017) can be expressed as:

$$P_{TRANS} = \left(P_{laser} + P_{driver} + P_{I/O}\right)_{TX} + \left(P_{PD} + P_{amp} + P_{I/O}\right)_{RX} \quad (4.24)$$

where $P_{laser}, P_{driver}, P_{I/O}, P_{PD}, and\ P_{amp}$are the consumed power by direct-modulated laser, electronics driving the laser, the electrical input/output interface, photodetector, and the trans-impedance and limiting amplifiers, respectively. This thesis assumes using of point to point transceivers (PtP) instead of point to multipoint since the loss of the PtP is a function of the distance and operating wavelength, i.e. the link losses in the PtP case is approximately as low as 6dB through a 20km network range (Yuh-Shyan et al., 2018). Therefore, the total consumed power in the C-RAN (P_{CRAN}) is evaluated by integrating the consumed power of the three main components of the network with consumed power through other functions (P_{others}) like AC-DC, DC-AC and cooling:

$$P_{CRAN} = \sum\left(P_{BBU} + P_{TRANS_B}\right) + \sum\left(P_{RRH} + P_{TRANS_R}\right) + P_{others} \quad (4.25)$$

where P_{TRANS_B} and P_{TRANS_R} exploits the consumed power of PtP transceivers connected at each BBU and RRH, respectively.

4.4.2 Parameterized and linear PM

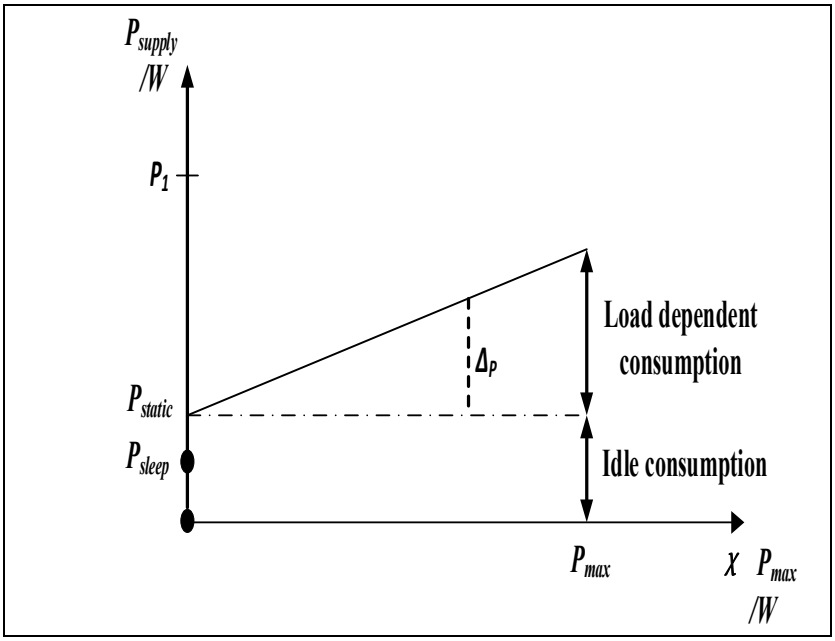

Figure 4.2 load-dependent power

According to (Alhumaima et al., 2016), if a base station consists of a BBU and a single RRH, the power used by the base station could be estimated as an affine function of the transmitted power. Then the consumed power consists of a load-dependent part, which increases linearly with the normalized traffic load of RRH n, x_n; $0 \leq x_n \leq 1$, from zero till the maximum value, $P_{n,max}$, and a static load-independent part, $P_{n,static}$. On the other hand, when base station n is not transmitting any signals, it is in an idle mode (sleep mode) with minimum power consumption ($P_{n,sleep}$). Then, the power consumed by a single BBU, serving a single RRH can be defined as:

$$P_C(x_n) = \begin{cases} P_{n,static} + x_n P_{n,max} & for\ 0 < x_n \\ P_{n,sleep} & for\ x_n = 0 \end{cases} \tag{4.26}$$

where x_n is normalized traffic load in the RRH. Then, the total consumed power in C-RAN is equivalent to the power consumed for all active BBUs and RRHs in the network:

$$P_C(\boldsymbol{x}) = \begin{cases} \sum_n (P_{n,static} + x_n P_{n,max}) & for \quad n : 0 < x_n \\ \sum_n P_{n,sleep} & for \quad n : x_n = 0 \end{cases} \tag{4.27}$$

where $\boldsymbol{x}$ is a vector of normalized traffic loads for active RRHs. The fundamental power model which presented in equation (4.26) is parameterized to discover the contribution of the different parameters. Moreover, Parameters that are supposed to be constant or having negligible effects are highlighted as well. The next approximations are introduced:

- The power consumption of the BBU, as well as the Radio Frequency (RF), are scaling linearly with the amount of bandwidth and the numbers of antennas (A).

$$P_{BBU} = L\left(\frac{\widehat{W}}{BW_{TOTAL}}\right) P_{BBU}^{PM} \tag{4.28}$$

$$P_{RF} = L\left(\frac{\widehat{W}}{BW_{TOTAL}}\right) P_{RF}^{PM} \tag{4.29}$$

Where P_{BBU}^{PM} and P_{RF}^{PM} are power consumption parameterized of BBU and RF, respectively

- In each RRH antenna unit, there is a power amplifier (PA), where its consumed power relies on the maximum transmission power for each antenna unit ($\frac{P_{max}}{L}$) and its efficiency (η_{PA}). The feeder losses (σ_{feed}) between the PA and the antenna can be neglected because PAs can be located near the antennas.
- The loss factors of DC-DC, AC-DC conversions, main supply units (Cho et al.), and cooling power consumption for the BBU pool are represented by $\sigma_{DC,POOL}$, $\sigma_{MS,POOL}$, and $\sigma_{COOL,POOL}$. For the RRHs, the factors of loss are interpreted by $\sigma_{DC,R}$ and $\sigma_{MS,R}$. Furthermore, the losses of the optical fibre between BBUs and RRHs are approximated by a loss factor $\sigma_{Optical}$.
- The optical transceivers power consumption scales linearly with the BBUs and RRHs number. If a single BBU consumed power that serving an RRH is:

$$P_C(x_n) = \frac{L\left(\frac{\widehat{W}}{BW_{TOTAL}}\right) P_{BBU}^{PM} + P_{TRANS_B}}{(1-\sigma_{DC,POOL})(1-\sigma_{MS,POOL})(1-\sigma_{COOL,POOL})} + \frac{L\left(\frac{\widehat{W}}{BW_{TOTAL}}\right) P_{RF}^{PM} + (\frac{P_{max}}{L.\eta_{PA}}) + P_{TRANS_R}}{(1-\sigma_{DC,R})(1-\sigma_{MS,R})(1-\sigma_{Optical})} \quad (4.30)$$

Thus, the last KPI to be considered is the power consumption in C-RAN. As introduced previously, the power consumption depends on the load in the network, as well as the number of active BBUs needed to accommodate this load. Therefore, it is necessary to run the network with efficient power usage without activating more BBUs than needed or without deactivating BBUs that consequently leads to increasing the number of blocking events and degrading the NP.

The KPI power consumption in C-RAN can be represented as:

$$KPI_P = \begin{cases} 1 & if\ P_C(\boldsymbol{x}) = P_{sleep} \\ \dfrac{1}{1+P_C(\boldsymbol{x})} & otherwise \end{cases} \quad (4.31)$$

The KPI_P is assumed to be one if the system is operating with minimum possible power consumption.

4.5 Objective function formulation

For NP maximization in the C-RAN, the whole KPIs presented in the previous sections are weighted individually to express the function of the NP. Then the main objective is to optimize the NP function (F_{NP}) defined by the following equation:

$$Max\ F_{\mathrm{NP}} = KPI_{BC}^{w_1} * KPI_{STD}^{w_2} * KPI_{interHO}^{w_3} * KPI_{intraHO}^{w_4} * KPI_{FHO}^{w_5} * KPI_P^{w_6} \quad (4.32)$$

Subject to:

$$\sum_{K} R_{ik}^{t+1} = 1 \ \forall i$$
$$\sum_{B} R_{ib}^{t+1} = 1 \ \forall i \tag{4.33}$$
$$\sum_{SoR_k} B_{UM_i,k} \geq B_{RBs} \ \forall i$$

where w_1, w_2, w_3, w_4, w_5 and w_6 express the predefined KPIs priority levels. The first constraint indicates that each RRH should be assigned to only one sector at a given time. Furthermore, the second constraint specifies that each sector should be linked to one BBU at a given time. Finally, the last constraint shows that the number of scheduled RBs for the UEs served by a given sector should not exceed the total number of RBs defined by the network to that sector. After introducing the definition of the first and second constraints, the size of the search space is reduced from 2^{NK} into K^N. The best RRH-Sector-BBU mapping can be identified by searching the whole search space regarding all applicable RRH exhaustively to sector association. When the RRHs and sectors number increases, the possible RRH-Sector-BBU mapping solutions exponentially increase as well. Hence, the execution time of the algorithm exponentially increases too. Therefore, several algorithms are introduced in the following chapter for solving the RRH-Sector-BBU mapping as an optimization problem.

4.6 Illustrative example

Figure 4.3 and Table 4.1 shows how the proposed KPIs are calculated. Let us assume 10 RRHs, and each RRH is serving a specific number of users represented by the number inside each RRH hexagonal cell. Furthermore, let us assume that the maximum capacity of each sector is 25 users. For the sake of simplicity, let us assume that the KPI for the power equals to 0.01 and the weights for standard deviation, number of forced handover blocked UEs, number of inter handovers, number of intra handovers, number of forced handovers, and power consumption are 0.2, 0.4, 0.1, 0.1, 0.1, and 0.1, respectively. Note that the KPI for forced handover is one because it depends on the difference between the configuration at time *t* and *t+1,* and in this scenario, there is no configuration at time *t+1*.

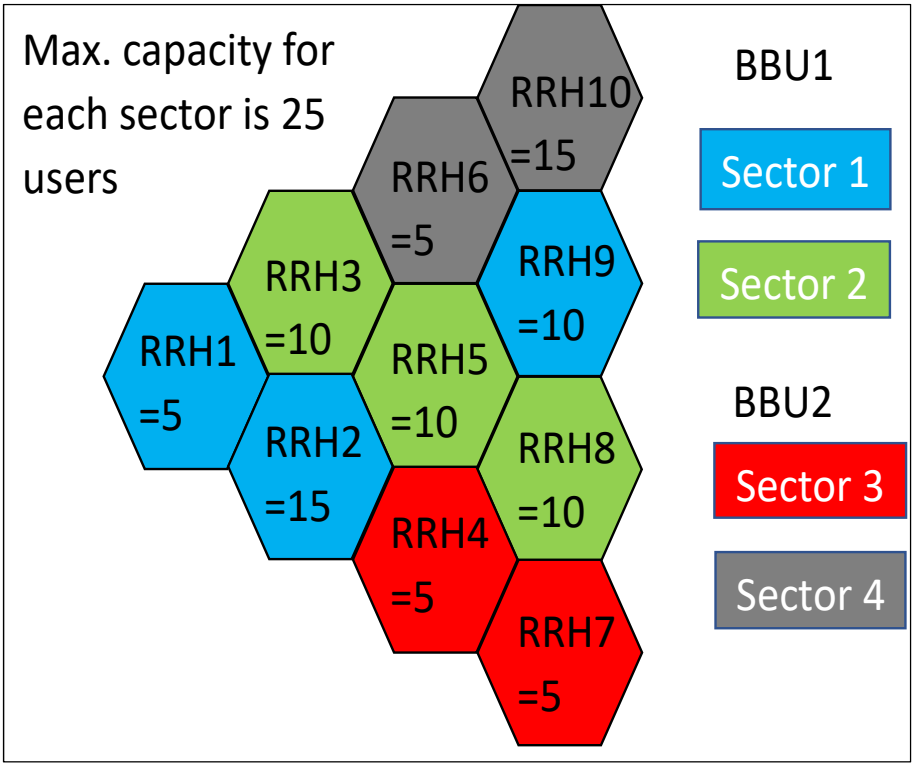

Figure 4.3 illustrative scenario for KPIs calculation

Table 4.1 KPIs calculation

KPI_{ψ}	$\left[\sqrt{\frac{(30-22.5)^2+(30-22.5)^2+(10-22.5)^2+(20-22.5)^2}{4-1}}\right]_{d=1}^{-1}$ $= [1+9.5743]_{d=1}^{-1} = 0.09457$
KPI_{BC}	$[1+(30-25)+(30-25)+(10-25)+(20-25)]_{d=1}^{-1}$ $=[1+(5)+(5)+(-5)+(-25)]_{d=1}^{-1} = [1+10]_{d=1}^{-1} = 0.0909$

Table 4.1 KPIs calculation (continued)

KPI_{inter}	$\left[(1)+0+0+\frac{5*1}{2}+0+\frac{5*1}{2}+\frac{5*1}{3}+0+0+\frac{5*1}{3}+0+0+15 * 1+0+\frac{15*1}{\sqrt{3}}+\frac{15*1}{2}+0+0+\frac{15*1}{\sqrt{7}}+0+0+\frac{10*1}{\sqrt{3}}+0+10*1+\frac{10*1}{\sqrt{7}}+0+0+\frac{10*1}{2}+\cdots\right]_{d=1}^{-1}$ $= [1+275.32]_{d=1}^{-1}= 0.003621$
KPI_{intra}	$\left[(1)+0+5*(1-0)+0+\frac{5*(1-0)}{\sqrt{3}}+0+0+\frac{5*(1-0)}{\sqrt{7}}+0+0+0+15*(1-0)+0+15*(1-0)+0+0+\frac{15*(1-0)}{\sqrt{3}}+0+0+10*(1-0)+10*(1-0)+0+0+0+0+0+\frac{10*(1-0)}{\sqrt{3}}+0+0+0+0+0+\frac{5*(1-0)}{2}+0+0+0+\frac{5*(1-0)}{\sqrt{7}}+\cdots\right]_{d=1}^{-1}$ $= [1+153.32]_{d=1}^{-1}= 0.006522$
$KPI_{f_{HO}}$	1
KPI_P	Calculated from subsection 4.5
F_{NP}	$0.09457^{0.2} * 0.0909^{0.4} * 0.0036213^{0.1} * 0.006522^{0.1} * 1^{0.1} * 0.01^{0.1}=$ 0.085

4.7 Chapter Summary

This chapter addresses the issue of inappropriate mapping between the RRH-Sector-BBU that can cause load imbalances in the C-RAN. This is because the overloaded Sectors can face resource shortage that degrades QoS when UEs try to connect to those Sectors, although there are nearby under-loaded Sectors that could serve these UEs.

The approach aims also to maximize other network performance metrics, which is done based on various KPIs. The proposed method introduces novel KPIs which are the standard deviation of the sector loads, number of blocked users, number of handovers, and the power consumption. The next chapter introduces the evolutionary optimization algorithms that are used to find the optimum RRH-Sector-BBU mapping.

CHAPTER 5

EVOLUTIONARY ALGORITHMS FOR RRH-SECTOR-BBU MAPPING

5.1 Introduction

In this chapter, we presented four selected optimization algorithms for finding the optimal RRH-Sector-BBU mappings: Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Bee Colony Optimization (BCO) and Cuckoo search (CUCO). All of them belong to the group of swarm-based optimization algorithms, defining feature being that they use a population of solutions for each iteration, rather than using one. If there is one optimal solution, it is expected that the members of the population will converge to it. If more than one local minimum is there, the population can capture them and keep them in the final result. All the algorithms mentioned above are loosely based on nature's methods to conduct the search and achieve an optimal solution.

5.2 Bee colony (BCO)

Bee Colony Optimization algorithm is also a swarm-based technique since it uses several points (bees) N^{Bee} simultaneously to search for the optimal solution. Each point here represents a realization of the RRH-Sector-BBU association vector R. When the number of RRHs is *N*, each such solution is a point in an *N*-dimensional space. The algorithm seeks to maximize the NP function by finding the optimal location in this *N*-dimensional space. Here we are going to outline the steps that the algorithm takes to perform the optimization.

Step 1*:* The algorithm generates a random initial population of solutions R^0(the upper index indicates the iteration number; zero for the initial conditions), consisting of N^{Bee} (number of bees) points, each represented by the *N*-dimensional vector R_i. The elements of the set are $R_i^0 = R_i$, where $1 \leq i \leq \mathrm{N}^{\mathrm{Bee}}$.

Step 2*:* NP values are calculated for each point using the supplied fitness function F: $f^0 = F(R^0)$ and the solution set R^0 is sorted accordingly to the solution fitness $f_i^0 = F(R_i)$.

Step 3: Number N^{elite} of best solutions from R^0 form the elite set E^0 of iteration "0": $E_0^i = R_i$, if $1 \le i \le N^{elite}$. Next, the number N^{sel} form the set of the selected sites S^0: $S_0^i = R_i$, where $N^{elite} + 1 \le i \le N^{elite} + N^{sel}$.

Step 4: The points which do not belong to either the elite set or the selected site set are discarded and reassigned for the local search near the solutions of the elite set and selected site set. For the elite set, we pick $R_j^{new}(E_i)$, where $1 \le j \le n^{elite}$, $1 \le i \le N^{elite}$, and condition $d(E_i, R_j) < d_{max}$ is held. Here d is the distance between two points R_i and R_j; $d(R_i, R_j)$ is defined as the number of non-coinciding vector components. That means for each solution from the elite set E_i, n^{elite} of random solutions are placed in its vicinity $d < d_{max}$.

Step 5: Similarly to step 4, n^{sel} of solutions are placed within a certain distance of each solution in the selected site set S^0(here $n^{elite} > n^{sel}$).

Step 6: NP is calculated for each newly assigned point of $R_j^{new}(E_i)$ and $R_j^{new}(S_i)$: $f^{new}(E_i) = F\left(R_j^{new}(E_i)\right)$ and $f^{new}(S_i) = F\left(R_j^{new}(S_i)\right)$ respectively.

Step 7: For each E_i and S_i, one best solution from the set $R_j^{new}(E_i)$, $R_j^{new}(S_i)$ respectively is selected. For all j values:

$$E_i = \begin{cases} R_j^{new}(E_i) & \text{if } F(E_i) < F(R_j^{new}(E_i)) \\ E_i & \text{if } F(E_i) > F(R_j^{new}(E_i)) \end{cases} \tag{5.1}$$

The same goes for all S_i:

$$S_i = \begin{cases} R_j^{new}(S_i) & \text{if } F(S_i) < F(R_j^{new}(S_i)) \\ S_i & \text{if } F(S_i) > F(R_j^{new}(S_i)) \end{cases} \quad (5.2)$$

Step 8: N^{elite} of the best solutions form the final sets E^o and S^o are assigned to the elite set E^0.

Step 9*:* The solutions that don't belong to E^0 are randomly assigned, and, together with the E^0 set they form the population of solutions for the next iteration R^1.

The procedure is repeated until the desired result is reached. As we can see, the algorithm combines global search, when new random RRH-Sector-BBU mappings are generated with subsequent local search. The local search takes place when the best mappings are mutated, assigning only a few RRHs to different sectors of the BBUs, thus placing the solution in the vicinity of the elite and selected solutions.

5.3 Particle swarm optimization (PSO)

Particle swarm optimization (PSO) also uses a set of particles called 'swarm' to engage optimization problems. Similarly to BCO, it performs both local and global optimization; however, with a different approach. In PSO, each particle 'remembers' its best-encountered position R_{Pbest}, and also the global best position encountered by the swarm R_{Gbest}. Apart from its position, the particle also has velocity, consisting of three components. Two components point to R_{Gbest} and R_{Gbest} from the current particle position, respectively. Their amplitudes are random so that the particle can be accelerated either more towards R_{Pbest}, or R_{Gbest}. The third velocity component retains the direction of the particle velocity from the previous iteration.

The search space of the C-RAN optimization is $N \times$ K, where N is the number of RRHs, and K is the number of sectors available. The solution R_i is a realization of the RRH-Sector-BBU association vector and has N components.

Here we will outline the steps that the algorithm takes to initialize and perform the optimization.

Step 1*:* The algorithm generates a random initial population of solutions R^0(the upper index indicates the iteration number; zero for the initial conditions), consisting of N^{swarm} (swarm size) points, each represented by the *N*-dimensional vector R_i. The elements of the set are $R_i^0 = R_i$, where $1 \leq i \leq N^{swarm}$. Correspondingly, for each particle i a random velocity V_i^0 is assigned, comprising the set of velocities V^0. Similarly to R_i^0, V_i^0 is the N-dimensional vector.

Step 2*:* NP values are calculated for each point using the supplied fitness function F: $f^0 = F(R^0)$ and the solution set R^0 is sorted accordingly to the solution fitness $f_i^0 = F(R_i)$.
Each particle's best position is set equal to the particle position $x_{Pbest,i}^0 = R_i$ on the 0-*th* iteration. In all following iterations the best position of a particle is:

$$R_{Pbest,i}^I = \begin{cases} R_{Pbest,i}^{I-1} & \text{if } F(R_i^I) < F\left(x_{Pbest,i}^{I-1}\right) \\ R_i^I & \text{if } F(R_i^I) > F\left(x_{Pbest,i}^{I-1}\right) \end{cases} \tag{5.3}$$

Step 3*:* The global best position R_{Gbest} is updated with the $R_{i_{MAX}}^0$ corresponding to the maximal $f_{i_{MAX}}^0$ value on the 0-*th* iteration. On the next iterations:

$$R_{Gbest}^I = \begin{cases} R_{Gbest}^{I-1} & \text{if } F\left(R_{i_{MAX}}^I\right) < F(x_{Pbest}^{I-1}) \\ R_{i_{MAX}}^I & \text{if } F\left(R_{i_{MAX}}^I\right) > F(x_{Pbest}^{I-1}) \end{cases} \tag{5.4}$$

Step 4*:* Velocity is updated according to:

$$V_i^1 = \delta V_i^0 + \tau_1 \rho_1 \left(R_{Pbest,i}^0 - R_i^0\right) + \tau_2 \rho_2 (R_{Gbest}^0 - R_i^0) \tag{5.5}$$

where τ_1, τ_2 and δ are free parameters, and $\rho_1$and ρ_2 are random numbers in the interval [0..1]. The first term retains particle velocity from the previous iteration, the second term represents acceleration towards the best position of the particle, and the third represents the acceleration towards the global maximum. As the acceleration is proportional to the particle distance from

the respective best positions, the particle essentially experiences an elastic force, drawing it towards the best positions.

Step 5*:* Particle position is updated according to its velocity:

$$R_i^1 = R_i^0 + V_i^1 \tag{5.6}$$

Steps 2-5 are repeated until the desired NP function value is reached.

5.4 Genetic algorithm (GA)

The genetic algorithm also uses a set of solutions simultaneously to perform optimization. Its idea is based on the rules of genetic heritage, mutation and natural selection. Due to the discrete nature of the genome, the GA is useful for discrete-valued optimization problems.
Let us assume that an N-dimensional vector represents each solution. In each new iteration, several vectors from the set, corresponding to the best cost function estimates, are retained as an elite group. Part of the vectors are the result of the crossover of the previous generation, i.e., each vector of the new generation contains parts of the vectors from the previous generation. Finally, a fraction of vectors is mutated, randomly changing their components.

In our case each solution R_i is a realization of a RRH-Sector-BBU association vector and has N components. Each component is a natural number spanning the number of sectors available K. We will outline the steps that the algorithm takes to initialize and perform the optimization.

Step 1*:* The algorithm generates the initial population of solutions R^0(the upper index indicates the iteration number; zero for the initial conditions), consisting of N^{pop} elements $R_i^0 = R_i$, where $1 \leq i \leq N^{pop}$.

Step 2: NP values are calculated for each point using the supplied fitness function F: $f^0 = F(R^0)$ and the solution set R^0 is sorted accordingly to the solution fitness $f_i^0 = F(R_i)$.

Step 3: Number N^{elite} of best solutions from R^0 form the elite set E^1 of iteration "1": $E_i^1 = R_i$, if $1 \leq i \leq N^{\text{elite}}$, and, as such, they are assigned to the solution set of the next iteration R^1 without modification.

Step 4: A selected fraction γ of the solution set R^1 is generated by crossovers of the vectors from the solution set R^0. A vector for the new iteration is generated by concatenating parts of the vectors from the previous iteration.

Step 5: The rest of the vectors of the solution set R^1 are generating by applying 'mutations,' i.e., adding variables with random distribution to the components of vectors from the set R^0. Usually, Gaussian distribution random variables are added; the standard deviation is usually changed in the process of optimization.

Steps 2-5 are repeated until the desired result is reached.

5.5 Cuckoo optimization algorithm (CUCO)

The cuckoo optimization algorithm is another type of nature-based optimization strategies. It attempts to mimic the behaviour of cuckoo birds, also known as brood parasites.
Cuckoos lay their eggs into the nests of other birds and thus minimize their effort in raising their offspring. Here we will list the main steps of the cuckoo optimization.
We start by introducing two sets of *R*-vectors, the positions of the cuckoo birds and the positions of their eggs, and for *i*-th iteration denote them R_C^i and R_E^i respectively.

Step 1: The initial distribution of the cuckoo positions R_C^0 is generated randomly, it contains N^{CO} elements.

Step 2: The number of eggs n_j^{Egg} for each cuckoo j is generated randomly from a uniform distribution, allowed range is from $n_{min}^{Egg} = 5$ to $n_{max}^{Egg} = 20$ eggs per cuckoo.

Step 3*:* Each cuckoo j ($j = 1 \ldots N^{CO}$) lays eggs within a certain radius (egg-laying radius ELR). The radius depends on the number of eggs produced n_j:

$$ELR_j = \alpha \frac{n_j}{n_{max}} (u - l) \tag{5.7}$$

where u and l are upper and lower bounds of the parameter space, and α determines the radius scaling. We define the distance between two *R*-vectors as a number of different components between them.

Step 4*:* For each cuckoo j a number n_j^{Egg} solutions are randomly placed within its ELR_j; they comprise the set R_E^i.

Step 5*:* NP values are calculated for each solution of the set R_E^i by use of the supplied fitness function F: $f^i = F(R_E^i)$, and the solution set R_E^i is sorted accordingly to the solution fitness $f_i^0 = F\left(R_E^i\right)$.

Step 6*:* To maintain the number of solutions stable, N^{CO} first solutions of the sorted set R_E^i are retained, and the rest are discarded.

Step 7*:* The points of the updated set R_E^i are clustered into 3 clusters Cl using a k-means algorithm according to the distance.

Step 8*:* The average fitness values of the solutions of each cluster Cl are evaluated:

$$\overline{f_{Cl}} = \frac{1}{n_{Cl}} \sum_i f_{i \subset Cl} \tag{5.8}$$

where n_{Cl} is the total number of cluster elements. The cluster with the highest fitness value is selected as the best. The centroid of the best cluster c_{Best} is established.

Step 9*:* For all solutions R_i which do not belong to the best cluster, the distance d_i to the centroid c_{Best} is determined accordingly to the chosen distance metric.

Step 10*:* The migration vector MV_i is determined by multiplying respective d_i with a randomly generated $\lambda \subset 0\ldots1$ to randomize the migration distance, and by randomizing the angle of migration. The angle ω between λd_i and MV_i is restricted to $\frac{\pi}{6}$. The migration vector is added to the respective position of the cuckoo, thus the starting population of the next iteration R_C^1 is determined.

The steps 2-10 are repeated until the desired outcome is reached or the maximum number of repetitions takes place.

5.6 Chapter Summary

In this chapter, we presented four evolutionary algorithms: Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Bee Colony Optimization (BCO) and Cuckoo search (CUCO). All of them belong to the group of swarm-based optimization algorithms, and they are used to find the optimum RRH-Sector-BBU mapping to balance the load across the network. All the algorithms are explained by steps to show how they can search for the optimum solution. The next chapter presents the performance evaluation of the models proposed in this thesis. The first part presents and analysis the RRH-Sector selection results, while the second part presents and analysis the RRH-Sector-BBU load balancing results.

CHAPTER 6

PERFORMANCE EVALUATION OF THE PROPOSED MODELS

6.1 Introduction

The main contributions in our work are to overcome the high complexity limitation of the previous works and develop the RRH-Sector selection MDP based model which considers the user and operator utilities together that is implementable in real systems. And, to present power efficient load balancing through dynamic RRH-Sector-BBU mapping. In this chapter, we introduce how we reach these contributions by firstly presenting the tested C-RAN network scenarios used in the simulations. Then, we compare the RRH-Sector selection models' performance in terms of operator's reward, users' average data rate, and users' connection demands blocking probability. The comparison is between the MDP based model with reward penalty (MDP-P), the MDP based model without the penalty (MDP-N), and the benchmark RSS model. Then we compare solutions for the RRH-Sector-BBU mapping optimization obtained by the evolutionary algorithms (BCO, CUCO, GA, and PSO) with the optimal solution obtained by exhaustive search (Ahmed et al.). The comparison includes performance sensitivity to the weights associated with the considered KPIs: load standard deviation, forced handover blocked users, handovers, and BBU power consumption.

6.2 Tested scenarios

The network parameters used in the simulated scenarios are presented in Table 6.1. The network is composed of 37 RRHs, where each RRH is serving a group of UEs. The user locations are generated independently for each connection and the connection arrivals form the Poisson process while the holding times are exponentially distributed. The 37 RRHS are grouped in 9 sectors, which are served by 3 BBUs. In general, C-RAN could operate with 1.5, 5, 10, 15, or 20 MHZ bandwidth. Here we assume that the considered C-RAN operates with

20 MHZ bandwidth, divided into 100 RBs according to the standard where each UE is going to get a minimum of 4 RBs when the sector reaches its maximum capacity.
Assume the hexagonal cell layouts for ease of result analysis and presentation, as shown in Figure 6.9 Initial mapping at time period t and mapping solutions at t+1. Table 6.2 shows the selected user data connection parameters.

Table 6.1 Network parameters

Parameters	**Values**
Number of RRHs	37
Number of Sectors	9
Number of BBUs	3
System bandwidth	20 MHz
Transmission power	36 dBm
Scheduler	FD-RR
Antenna mode	Isotropic
Inter-site distance (ISD)	500m
Transmission scheme	SISO
c_{RBG}	3
Fading	Standard deviation 4 dB, log-normal
Noise spectral density per Hz	-174dBm

Table 6.2 User data parameters

Parameter	**Value**
Packet request interval	5ms
Packet length	536 Bytes
Data size	2680 Bytes per request

The load offered to a sector *k* as defined earlier is:

$$A_k = \sum_{j \in J} \frac{\lambda_j}{\mu_j} \tag{6.1}$$

where λ_j is the connection demand arrival rate of class j and μ_j is the service rate of class j, where $j \in J$. Note that each sector has the same maximum capacity. Moreover, the sector nominal load ($\hat{A}_k$) is defined as the load which gives QoS with approximately 2% of blocking probability. Then, the normalized sector load (ϱ_k) is defined as the ratio between the offered load to a sector (A_k) to the nominal sector load ($\hat{A}_k$):

$$\varrho_k = \frac{A_k}{\hat{A}_k} \tag{6.2}$$

Thus, the normalized network load ϱ could be defined as:

$$\varrho = \frac{\sum_k A_k}{\sum_k \hat{A}_k} \tag{6.3}$$

The user utility calculation for the selected criteria is based on parameter values presented in Table 6.3 that indicates the minimum, mean, and maximum values demanded by the user and the associated weights for each criterion. The calculations of criteria values offered by the sectors are based on models presented in subsection 3.6. Then, the calculation of the utility for each criterion uses values from Table 6.3 and equation (3.9) from subsection 3.3. To assign the weights to each criterion, assume the following order of decreasing criteria importance: data rate, RSS, and the Service cost for the user. Also, to avoid neglecting some criteria, assume that the weight for any criterion must be below 0.5. The resulted constraints are as follows:

$$\begin{gathered} w_{dr} + w_{RSS} + w_{cs} = 1 \; Subject \; to: \\ w_{dr} > w_{RSS} > w_{cs} \\ w_i < 0.5, w_i \in \{w_{dr}, w_{RSS}, w_{cs}\} \end{gathered} \tag{6.4}$$

where w_{dr}, w_{RSS} and w_{cs} are the weights for data rate, RSS, and Service cost for the user, respectively.

Table 6.3 User requested criteria values and weights

Criterion	**Requested values from user: min, mean, max/weight**	
	Class 1 user	Class 2 user
Data rate (Mbps)	2, 3, 4 / 0.4	0.512, 1, 2 /0.5
RSS (dBm)	-80, -70, -60/0.35	-100, -95, -90/ 0.3
Service cost for the user	0, 3, 6/ 0.25	0, 2, 4 / 0.2

For the MDP-P model, the chosen values of $\hat{k}$, $\hat{k}_1$, $\hat{k}_2$, $\mathcal{T}_1$ and $\mathcal{T}_2$ parameters for the reward penalty calculation are listed in Table 6.4. The values of arrival, service, and reward rates are listed in Table 6.5. Note that the parameters for sectors in Table 6.5 are used for MDP model verification, where initially the sectors have the same coverage area which means that each sector consists of the same number of RRHs.

Table 6.4 Constant values for utility calculation

Constant	**Value**	**Description**
$\hat{k}$	0.6	Reward coefficient for zero utility.
$\hat{k}_1$	0.5	Reward coefficient for tolerance zone.
$\hat{k}_2$	0.2	Reward coefficient for satisfaction zone.
$\mathcal{T}_1$	0.5	Utility threshold for tolerance zone.
$\mathcal{T}_2$	0.8	Utility threshold for satisfaction zone.

Table 6.5 Traffic parameters for sectors

Parameters	Values
Sector capacity	250
Arrival rate for class 1 user λ_1	16 to 285
Service rate for class 1 user μ_1	2
Reward rate for class 1 user r_1	7
Arrival rate for class 2 user λ_2	130 to 900
Service rate for class 2 user μ_2	5
Reward rate for class 2 user r_2	5
Nominal load	240

6.3 Performance analysis of the RRH-sector selection algorithm

Figure 6.1 presents the network operator average reward as a function of the network load. The average reward of MDP based model with penalty can be represented in two ways. The first one is the average reward which is achieved only from the reward parameters (that can be interpreted as the revenue) of the connections and indicated as MDP-P(r). And the second one is the average reward which presents the reward from the connection reduced by the penalty and indicates as MDP-P. In general, the reward parameter can be interpreted as a control parameter or as a monetary revenue. However, the average reward from the MDP-P model can be interpreted only as a reward and not revenue since the penalty does not have revenue meaning in this case and may be treated as a node selection optimization parameter. The operator reward is the same for all the models when the network is underloaded with $\varrho = 0.14$, since the network has sufficient resources to accept the connections regardless of the different selection algorithms. MDP-P(r) gives the highest average operator reward (that can be interpreted as the revenue) compared to the other algorithms. Morover, the operator reward increases when using MDP-P instead of RSS for each network load value. Note that the MDP-P reward decreases in the overload conditions (i.e., network load exceeds 1), this is caused by the degradation of QoS parameters that increases the penalty applied to the reward rate. On the

other hand the reward for the MDP-N model (that can be interpreted as the revenue) keeps increasing with the network load increase.

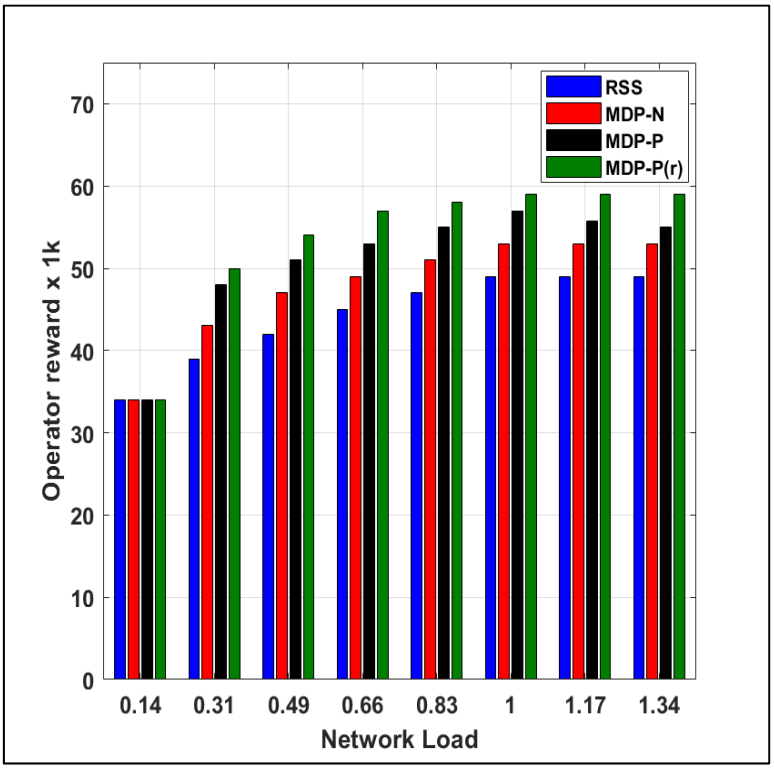

Figure 6.1 Operator revenue vs. Network load

The blocking probability is an important metric for communication networks since its high values can lead to user frustration. Morover, if the network blocks less connections, that means it is going to get more rewards. We estimate the blocking probability as the ratio of the number of rejected connection demands to the total number of demands. Figure 6.2 shows the estimated blocking probabilities as a function of the network load level. We can notice that the blocking probability increases slowly with the network load until the network load reaches values around 1, after which the increase becomes more significant. Both the MDP-P and MDP-N models have lower blocking probabilities than the RSS since the RSS model does not consider the sector loadings and depends only on the users' distances from the RRHs. Moreover, the MDP-P model provides smaller blocking probabilities compared to the MDP-N model. This is because the MDP-P algorithm tends to avoid already crowded sectors by considering the

penalty, while in the case of the MDP-N, the sector's choice is related to the average reward only. In addition, due to the MDP-P capability of getting the lowest blocking probability, it obtains the highest rewards compared to the MDP-N and RSS as shown in Figure 6.1.

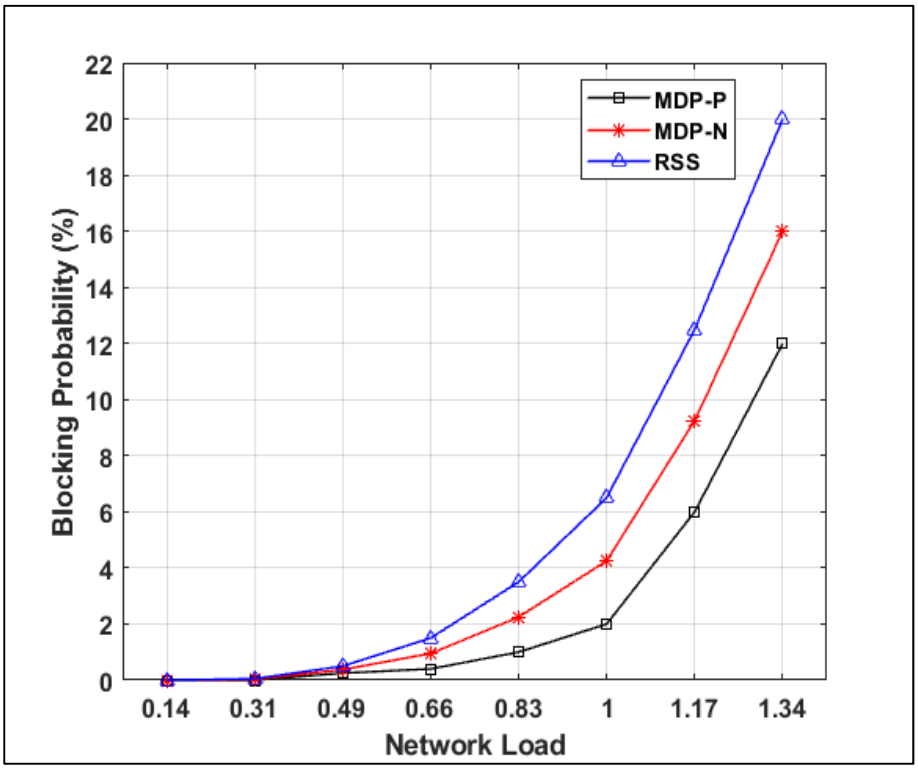

Figure 6.2 Users' blocking probability vs. Network load

InFigure 6.3, the average users' data rate vs. network load is shown. Since the data rate depends on the assigned number of RBs by the scheduler, therefore UE might obtain a different number of RBs depending on the number of connected UEs at a certain time. Thus, each UE is going to get a minimum of 4 RBs when the sector reaches its maximum capacity otherwise it could get more RBs. Here, the results indicate that when the network load increases, the average users' data rate decreases due to the network capacity limit and sharing the same amount of resources among more UEs. While the average users' data rate for the MDP-P, MDP-N, and RSS have the same value for the low load, for larger loads the MDP-P and MDP-N models provide larger average user' data rates when compared to the RSS model. This is because the selection in the RSS model depends on the users' distances from the RRHs. Hence, the UEs

are more likely to chose overloaded sectors that can provide fewer resources per UE. Furthermore, the MDP-P model provides higher users average data rate compared to the MDP-N. This is beacuse the MDP-P model considers the data rate in the user reward penalty as well as the average reward of the operator, while the MDP-N cares only about maximizing the average reward of the operator.

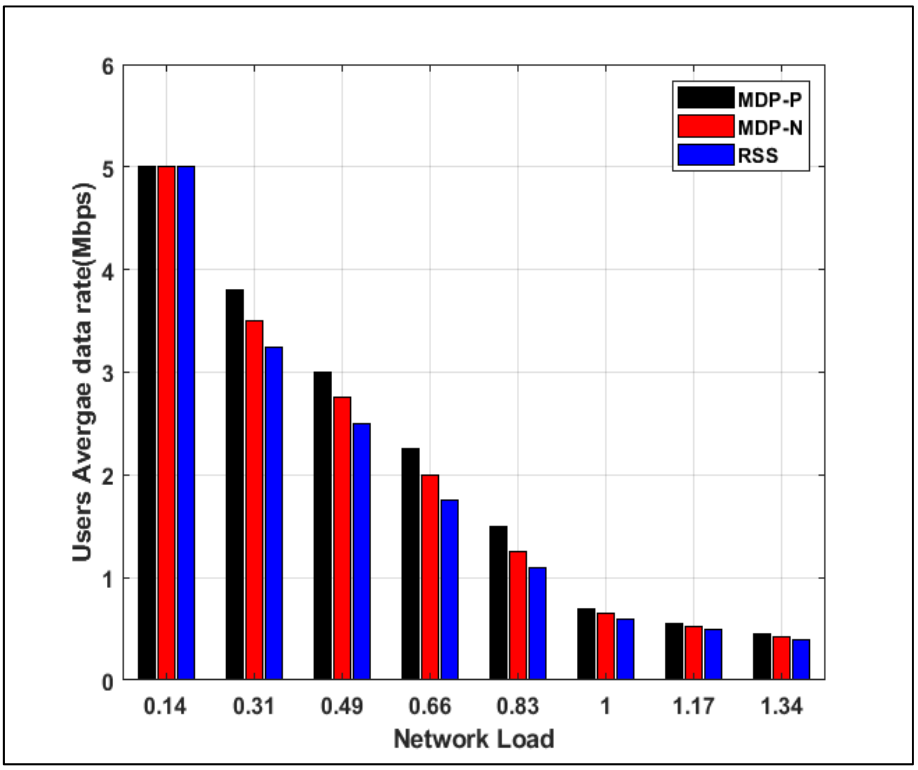

Figure 6.3 average users' data rate vs. network load

6.4 Performance analysis of the RRH-Sector-BBU allocation low load scenario

In this section, the performance of the proposed solutions for mapping between BBUs, RRHs, and Sectors, that balances the load and improves the NP at time period t+1 under a given network condition at time period t, is analyzed. In the first part of this section, we assume that network load $\varrho = 0.4$ and the weight for the forced handover blocking KPI is 0.5 and 0.1 for

each of the remaining KPIs. Other KPI weight value combinations are analyzed in the second part of this section.

We tested the BCO, PSO, CUCO, and GA algorithms for 20 scenarios with different initial RRH-Sector-BBU mappings. The solutions obtained from the evolutionary algorithms are also compared with optimal values obtained by exhaustive search (Ahmed et al.).

Figure 6.4 shows the NP function values (averages over 20 scenarios) as a function of the iteration number and the NP function value from the optimal EX solution. PSO, GA, BCO, and CUCO obtained 0.754, 0.754, 0.756, and 0.757, respectively, while the optimum value obtained from the optimal EX solution is 0.7574. The analysis of individual scenarios show that the optimum solution is reached by each evolutionary algorithm 18 times out of 20 scenarios. Note that the not optimal solutions can be obtained for different set of scenarios for each of the evolutionary algorithms.

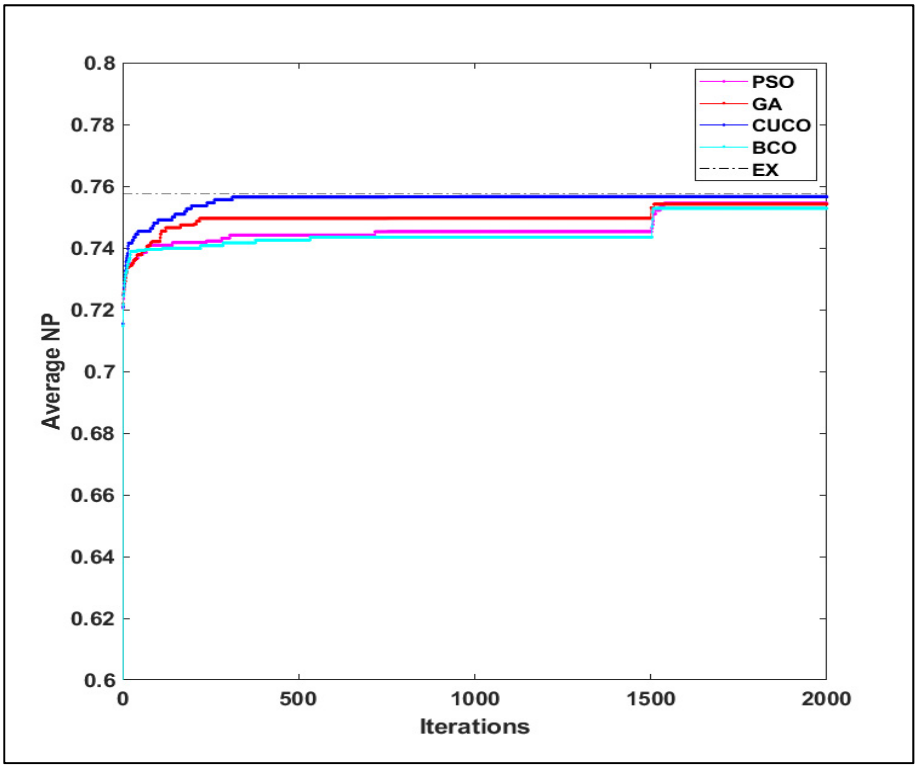

Figure 6.4 Average NP vs. Iterations for 37 RRH

Figure 6.5 represents the average number of forced handover blocked users over the 20 scenarios as a function of the iteration number. Note that all the evolutionary algorithms succeeded in reaching zero forced handover blocked users, as is the case in EX, in all the 20 scenarios.

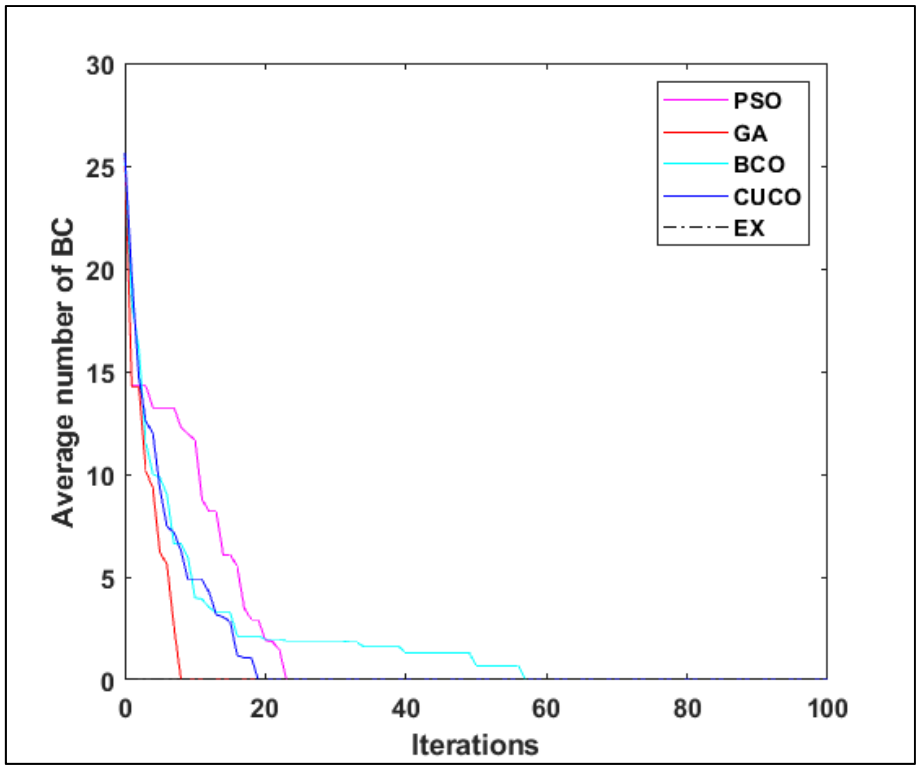

Figure 6.5 Average BC for 37 RRH

Figure 6.6 represents the average number of intra handovers as a function of the iteration number. BCO achieves the smallest value of average number of intra handovers equals to 3111. GA and PSO reach 3357 and 3445, respectively and CUCO obtains 3482, which is close to the average optimum solution provided by the EX that equals 3650. Note that the BCO, GA, and PSO algorithms provide a smaller number of average intra handovers compared to the EX; however, the KPIs of EX and CUCO indicate better values of the average NP function.

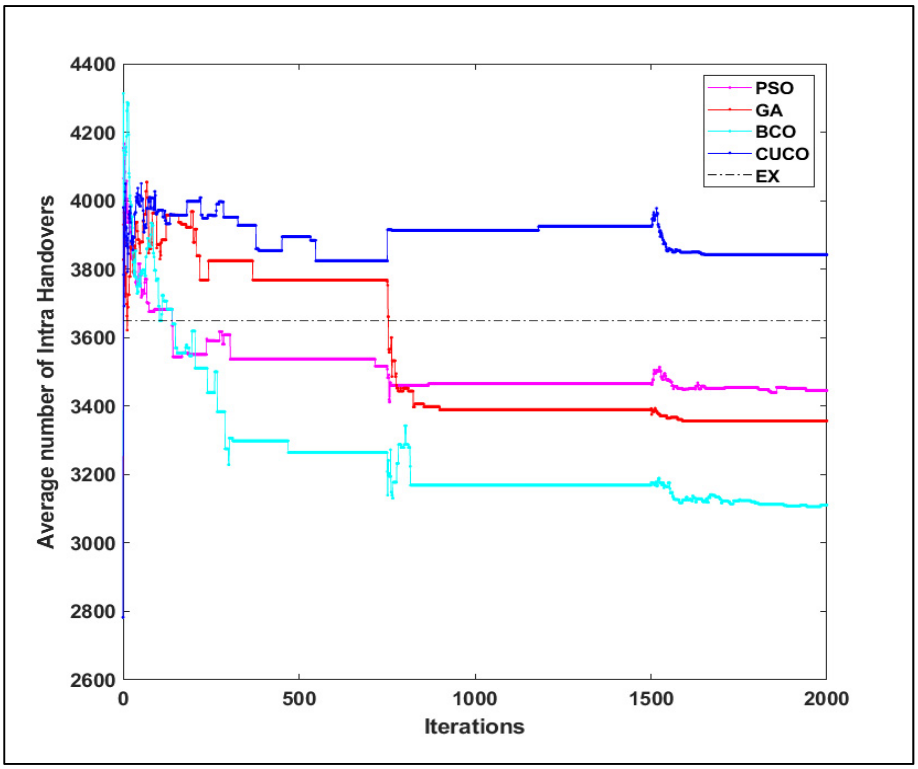

Figure 6.6 Average number of intra-handovers for 37 RRH

Figure 6.7 shows the average number of inter handovers as a function of the iteration number. CUCO achieves the smallest number of average inter handovers equals to 6441. Both GA and PSO reach 6674, and BCO obtains 6653. None of the evolutionary algorithms succeed to achieve the optimum average number of inter handovers provided by EX that equals to 6119.

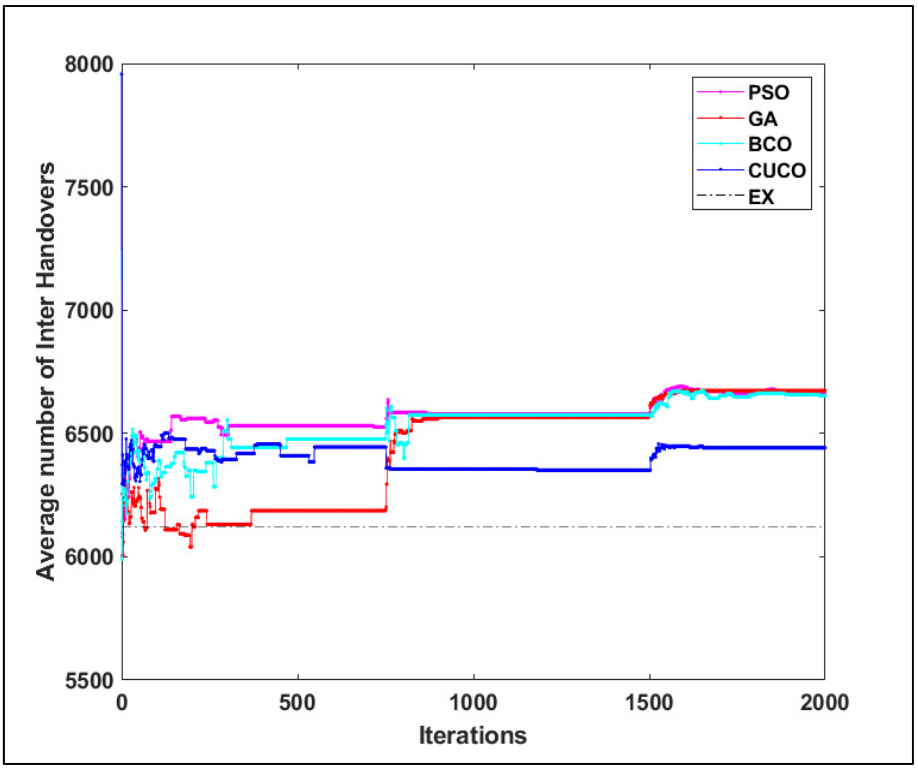

Figure 6.7 Average number of inter-handovers for 37 RRH

Figure 6.8 represents the average number of forced handovers as a function of the iteration number. CUCO achieves the smallest average number of forced handovers equal to 488 which is close to the average optimum solution provided by the EX that equals to 463, while BCO and PSO reach 520, and GA obtains 606.

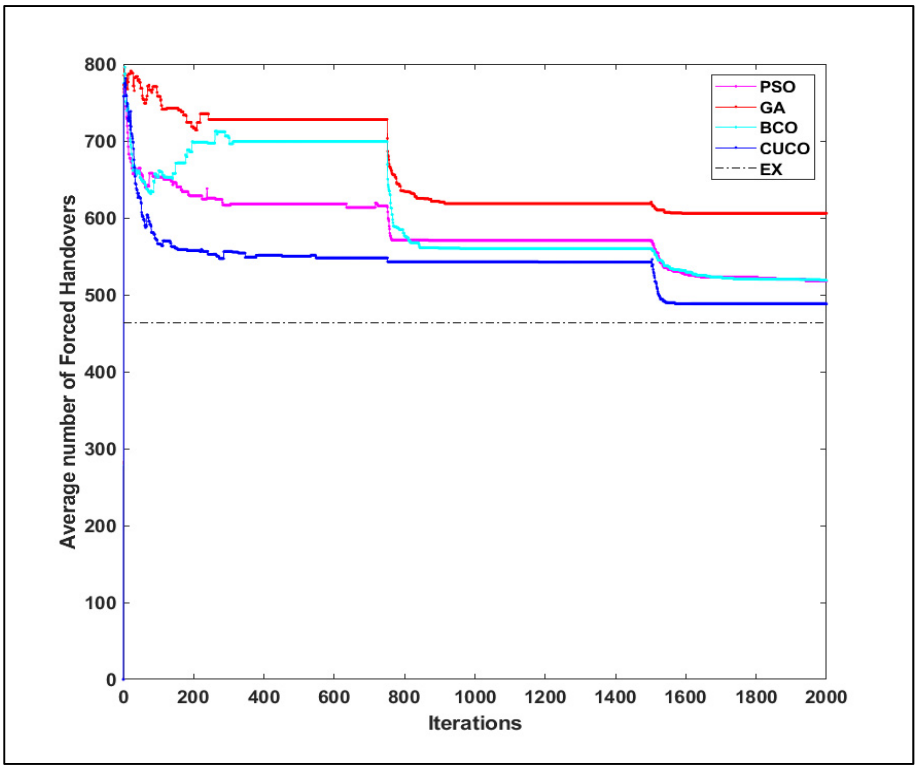

Figure 6.8 Average number of forced handovers for 37 RRH

Note that the difference between the average solutions numbers of the EX and the evolutionary algorithms presented in Figure 6.4Figure 6.8 comes from only 2 different scenarios out of the 20 scenarios for each evolutionary algorithm.

To give a better illustration and more details of the solutions provided by the BCO, PSO, CUCO, and GA algorithms, Figure 6.9 illustrates the solution topologies obtained by the EX and evolutionary algorithms in one selected scenario from the 20 scenarios. The selected scenario is the one that gives the maximum NP function value obtained by the EX out of the 20 scenarios. Figure 6.9 shows that in this scenario CUCO and BCO obtained the optimal RRH-Sector-BBU mapping, as given by EX, while the GA and PSO algorithms obtained suboptimal solutions. These GA and PSO results correspond to one of the two scenarios, out of the 20 scenarios, in which they failed to provide the optimum solution. Note that, some sectors are not fully continuous because as mentioned earlier the assigned weight for this case is 0.5 for the forced handover blocking KPI and 0.1 for each of the remaining KPIs. Thus, the evolutionary algorithms give the priority to optimise the forced handover blocking which in

return maximizes the NP function value. It should be noted that some other combinations of KPIs give a continuous sectors when the inter and intra handovers are assigned very high weights. Table 6.6 shows the performance metric values obtained for the selected scenario. The CUCO and BCO algorithms maximized the NP function as did the EX. PSO and GA succeed in reaching near optimum value, which constitutes approximately 98% of the optimum solution. Also note that the PSO and GA might reach the optimum solution in this selected run if we increase the population and swarm size (Srinivas et al., 1994; Y. Zhang et al., 2015). Note that, both PSO and GA are the first and second to converge, but they did not reach the optimum solution. On the other hand, BCO is the fastest one to converge and reach the optimum solution, while the CUCO search took more time to converge to the optimum solution.

Table 6.6 Computational results for a selected scenario out of the 20 scenarios (37 RRH)

KPIs vs. algo.	**NP Function value**	**Load standard deviation**	**Number of active BBUs**	**Number of forced handover blocked users**	**Number of Inter handovers**	**Number of Intra handovers**	**Number of Forced handovers**
Initial	0.109	70.1536	3	30	6119	3650	463
PSO	0.787	2.8868	2	0	6908	3535	560
GA	0.784	4.1833	2	0	6995	3757	505
BCO	0.799	2.2361	2	0	6895	2810	505
CUCO	0.799	2.2361	2	0	6895	2810	505
EX	0.799	2.2361	2	0	6895	2810	505

Table 6.7 Convergence results for a selected scenario out of the 20 scenarios (37 RRH)

KPIs vs. algo.	**Convergence (iterations)**	**Convergence (time in CPU sec.)**
Initial	NA	NA
PSO	754	25
GA	758	28
BCO	768	35
CUCO	1525	43
EX	9^{37}	43200

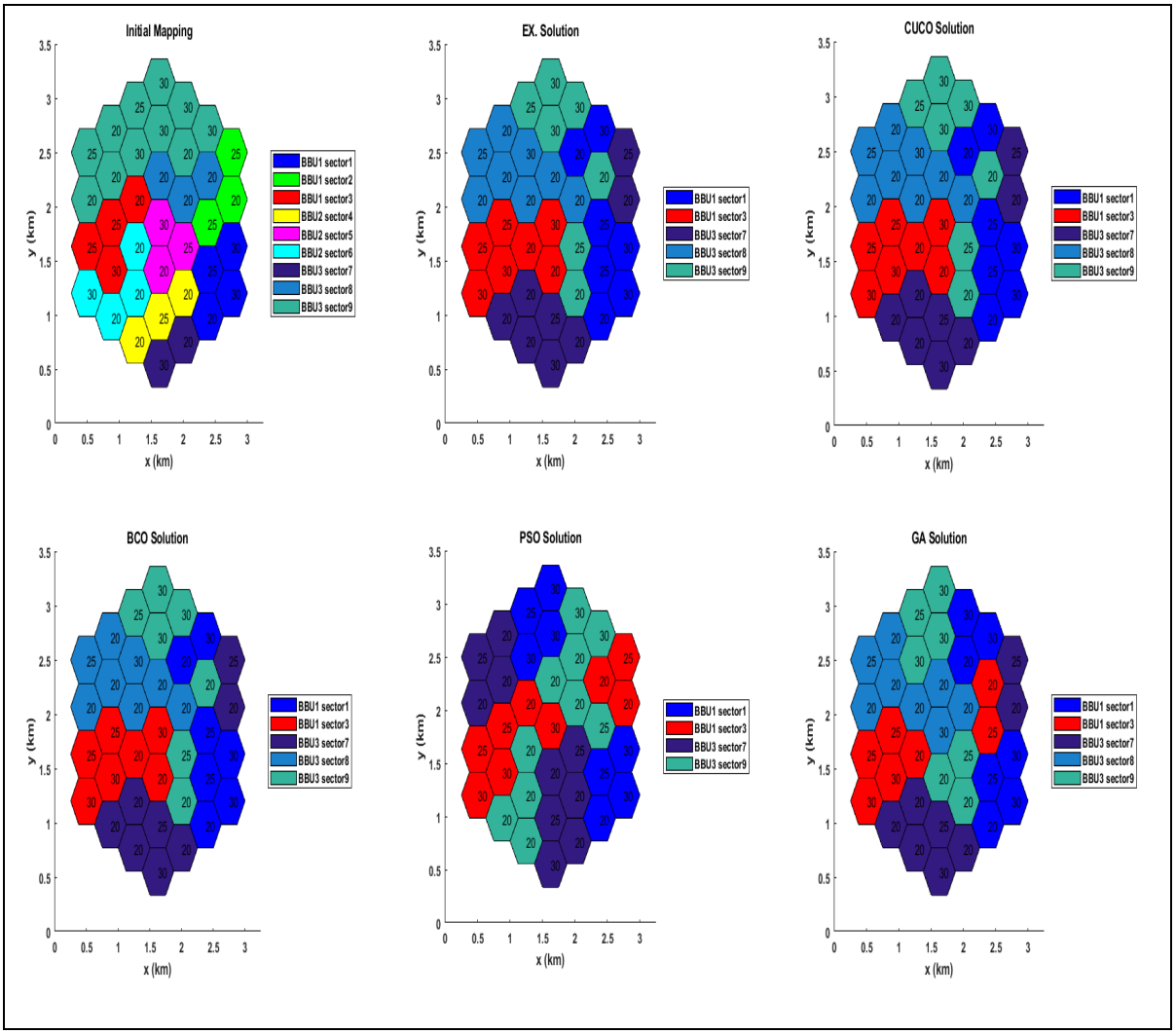

Figure 6.9 Initial mapping at time period t and mapping solutions at t+1

Figure 6.10 shows the load distribution among the sectors for the selected scenario. The load is fairly shared among the active sectors, reflecting the minimization of the load standard deviation by all the evolutionary algorithms compared to the initial mapping. The CUCO and BCO reach the optimum standard deviation obtained by the EX whereas the PSO provides standard deviation close to the optimum. Note that PSO is not allocating to sector 8 at all while allocating more UEs to the other sectors. On the other hand, the GA provides standard deviation better than the initial mapping but not close to the optimum obtained by the EX. Consequently, more UEs could be admitted in different sectors because the standard deviation

has been decreased by all the evolutionary algorithms. Figure 6.11 shows the BBU load for the optimized RRH-Sector-BBU mappings for the selected scenario. Note that the number of active BBUs required to support the current load of the network has been reduced to 2 (from 3) and that can save a considerable amount of power.

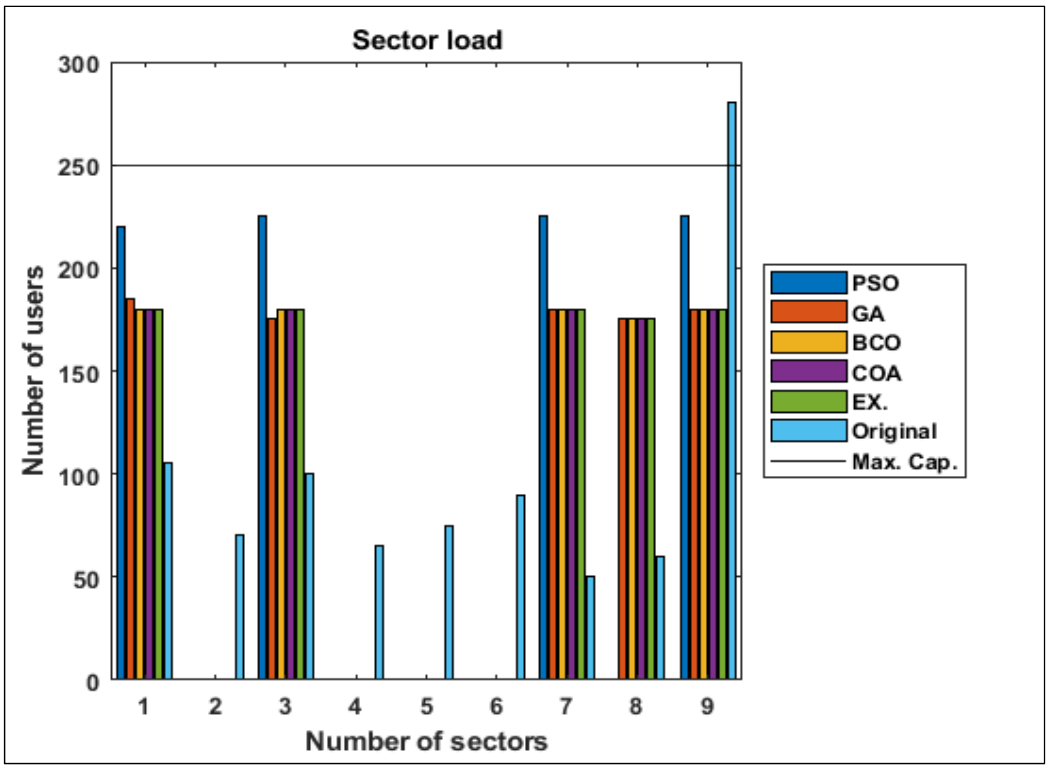

Figure 6.10 Sectors load for 37 RRH

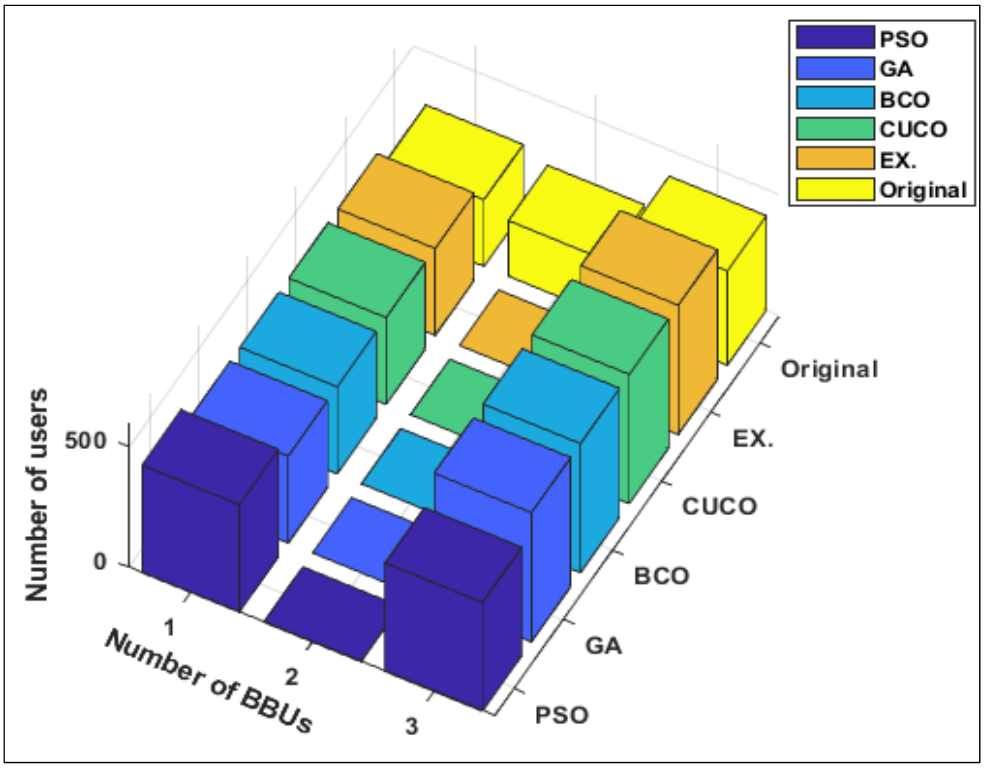

Figure 6.11 BBUs load for 37 RRH

Figure 6.12 shows the BBUs' utilization rates for each BBU for the selected scenario. BBU 2 is switched off by all the evolutionary algorithms in contrast to the initial configuration that uses three BBUs to operate the network. Moreover, the BBU utilization rate of the operating BBU 1, and BBU 3 for the GA, BCO, and CUCO reaches the optimum value obtained by the EX solution. On the other hand, the PSO BBU utilization rate of the operating BBU 1 is higher than the solution obtained by the EX; however, it gives smaller BBU utilization rate for the operating BBU 3.

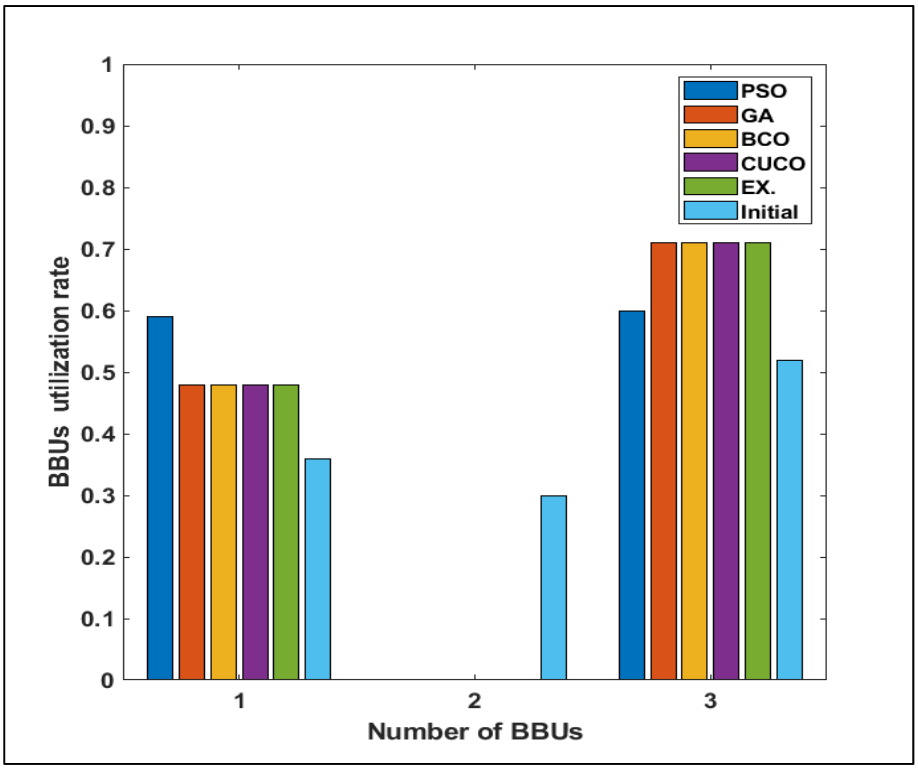

Figure 6.12 BBU resource utilization rate

Figure 6.13 shows the total power consumption comparison of all algorithms and the initial configuration for the selected scenario. Note that the total power consumption of the evolutionary algorithms (3000W) is reduced by 40% compared to the initial configuration (5000W).

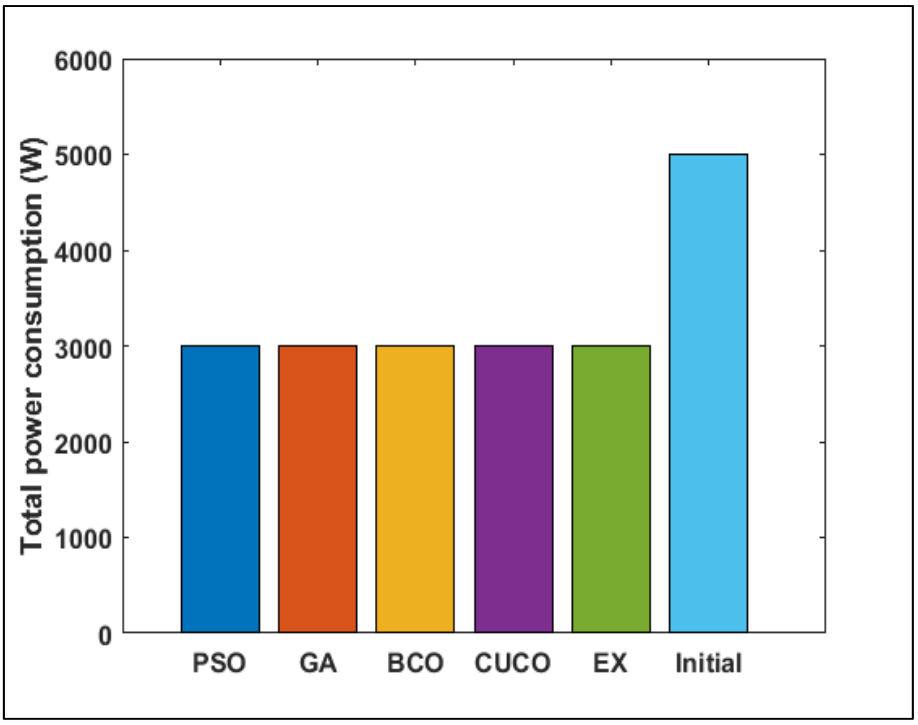

Figure 6.13 Total power consumption (W)

6.5 KPI scenario for weight sensitivity analysis

The previous results are calculated when the weight for forced handover blocking KPI is 0.5 and 0.1 for all remaining KPIs: the load standard deviation, power, forced handover, inter-handover, and intra-handover. In Table 6.8, we present the NP metrics sensitivity to changes in the KPIs' weight values. In particular, in each considered KPI scenario we set the weight of a specific KPI to 0.5 to dominate the other KPIs which are assigned weight of 0.1 each.

Table 6.8 shows the values of metrics related to each KPI and the NP function value for each KPI scenario obtained by the EX algorithm. When the BC weight is dominant, the number of forced handover blocked users and the standard deviation are minimized. In addition, the power consumption is minimized, which indicates that the number of operating BBUs equals to 2 instead of 3. However, there are large number of inter, intra and forced handovers because they are assigned low weights. On the other hand, the same result for the NP function value is reached in each of the following cases when one of the standard deviation, the inter handover,

intra handover, or power consumption has a dominant weight. This is because the EX solution favours increasing the number of forced handover blocked users and the number of forced handovers but it minimizes the other KPIs which in return gives the maximum NP function value. Finally, in case the forced handovers haves the dominant weight, the KPI of the number of forced handovers is minimized to 0. This is because the algorithms at time t+1 aim to keep the mapping of the RRH-Sector-BBU as the initial configuration at time t without changing it. Figure 6.14 demonstrate the NP function values obtain by all the evolutionary algorithms for each KPI scenario. The NP function value obtain by all the evolutionary algorithms is equal to the optimum value when one of the following KPIs which are; the standard deviation, the inter handover, intra handover, or power consumption has a dominant weight. Moreover, when the BC weight is dominant, the NP function values obtain by all the evolutionary algorithms are found to be equal. Furthermore, in case the forced handover has the dominant weight, the NP function values obtain by the CUCO and the BCO are equal to the optimum value. On the other hand, the PSO and GA obtain near optimum values.

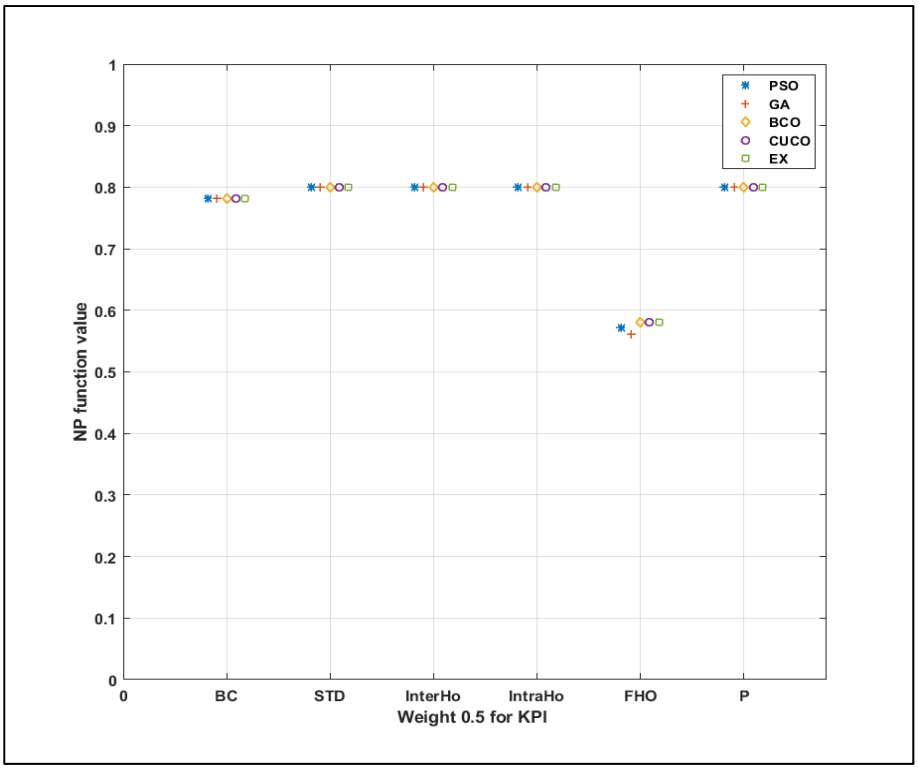

Figure 6.14 NP values for different weights combinations of the KPIs

In summary, the presented numerical results show that each evolutionary algorithm reaches the optimum solution in 18 out of 20 different initial mappings. In all the runs, the evolutionary algorithms succeeded to minimize the number of forced handover blocked users to 0 similar to the EX because it has the highest weight. Furthermore, the evolutionary algorithms give near optimum values for the other KPI matric (i.e., average number of intra handovers, average number of inter handovers, average number of forced handovers). Note that the CUCO, BCO, PSO gives closer values to the EX compared to the GA. This is due to that in PSO, the particles (RRH-Sector-BBU mappings) behaves as semiautonomous agents which are aware of each other's position status and decides to change their states (at each iteration) with respect to the best-observed particle position in the population. Similarly, the BCO uses a number of points (bees) simultaneously to search for the optimal solution. Each point represents a realization of the RRH-Sector-BBU mapping vector. On the other side, the chromosomes (RRH-Sector-BBU allocations) in GA are not agent-like and lacks the ability to sense the neighboring environment. Moreover, the results in the selected scenario indicate that both CUCO and BCO succeed to obtain the optimum NP function value while the PSO and GA reach 98% of the optimal value. Furthermore, the BCO is the fastest one to converge and reach the optimum solution, while the CUCO search took more time to converge to the optimum solution achieved by the EX. On the other hand, both PSO and GA converged quickly, but they did not reach the optimum solution. Finally, the results show that the KPIs are quite sensitive to the related weights so their choice should be considered carefully by the network operator according to its preferences.

Table 6.8 Optimum KPIs values for different combinations of KPIs' weight for the KPI scenario (37 RRH)

Weight 0.5 for KPI:	**Number of BC**	**Standard deviation**	**Number Inter HO**	**Number Intra HO**	**Number Forced HO**	**Power in (w)**	**Optimum NP function value**
BC	0	0	6745	3194	555	3000	0.7788
STD	645	0	0	0	615	3000	0.8003
interHO	645	0	0	0	615	3000	0.8003
intraHO	645	0	0	0	615	3000	0.8003
FHO	30	69.8	8000	3254	0	5000	0.5811
P	645	0	0	0	615	3000	0.8003

6.6 Performance analysis of the RRH-Sector-BBU mapping high load scenario

We consider a scenario with 61 RRHs and 4 BBUs. The network load is equal to $\varrho = 0.83$, and the KPI weights are assumed to be 0.6 for the blocked users, 0.1 for standard deviation, 0.1 for power, 0.1 for forced handover, and 0.05 for both inter-handover and intra-handover. The BCO, PSO, CUCO, and GA algorithms are repeated 20 times, with 20 different initial configurations. The solutions obtained from the evolutionary algorithms are compared with optimal values obtained by exhaustive search (ES).

Figure 6.15 shows the average NP for the evolutionary algorithms vs. the iteration number. The average NP value obtained by the ES solution, which represents the optimal solution, is 0.7918, while the PSO, GA, BCO, and CUCO obtained 0.711, 0.704, 0.707, and 0.727, respectively. Therefore, the evolutionary algorithms reach the optimum solution 18 times out of 20 runs.

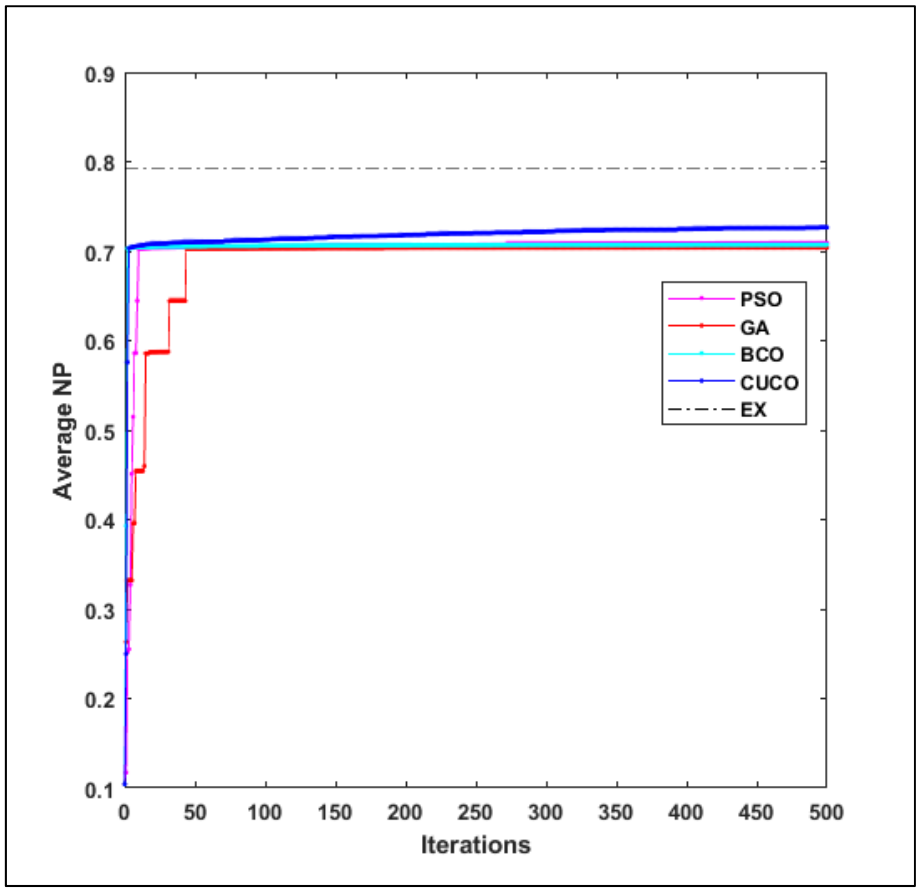

Figure 6.15 Average NP vs. Iterations for 61 RRH

Figure 6.16 represents the average number of forced handover blocked users during the 20 runs. All the evolutionary algorithms succeed to reach 0 blocked users similar to the ES. The only difference is the number of iterations that each algorithm takes to reach the 0 blocked users.

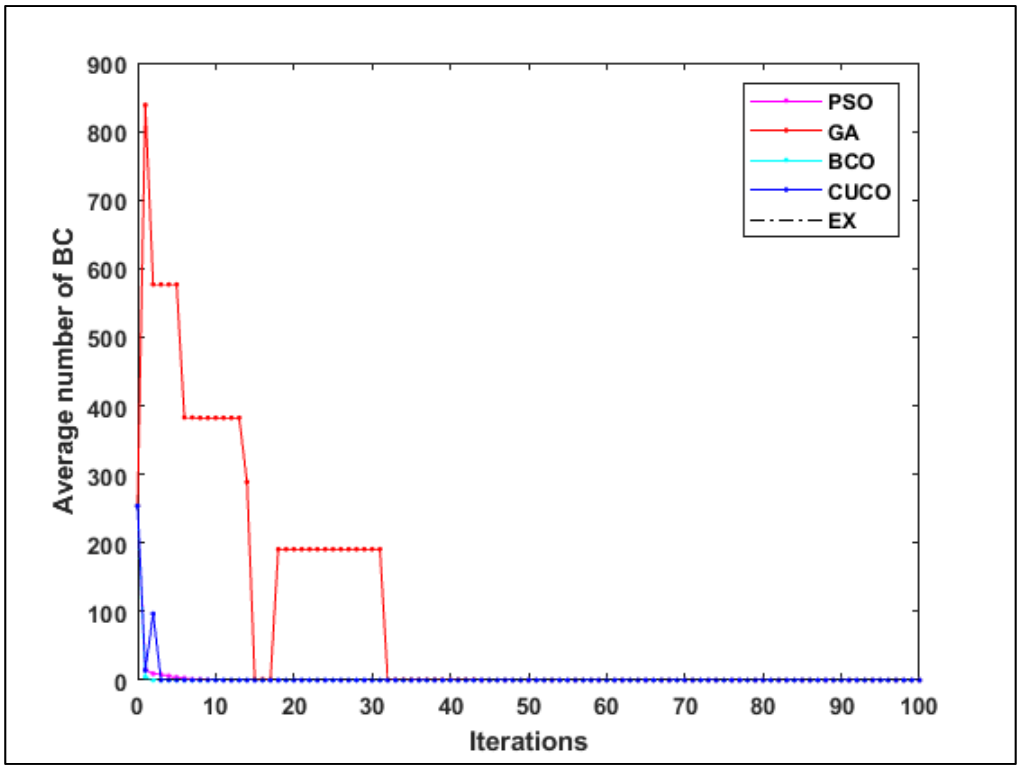

Figure 6.16 Average BC for 61 RRH

Figure 6.17, Figure 6.18 and Figure 6.19 represent the number of intra-BBU handovers, inter-BBU handovers, and forced handovers, respectively. The BCO, PSO, CUCO, and GA algorithms decrease the number of handovers required for the next RRH-Sector-BBU mapping; however, there is a small difference between their solution and the solution provided by the exhaustive search. This is due to the two runs that failed to reach the optimum solution throughout the 20 runs, as explained in Figure 6.15. It is clear that the CUCO algorithm is the one with the least number of intra handovers and forced handovers compared to the BCO, PSO, and GA algorithms. Furthermore, the BCO, PSO, and GA almost reach the same range for the intra and forced handovers. Moreover, the PSO algorithm is the one with the least number of inter handovers when compared to the BCO, CUCO, and GA algorithms.

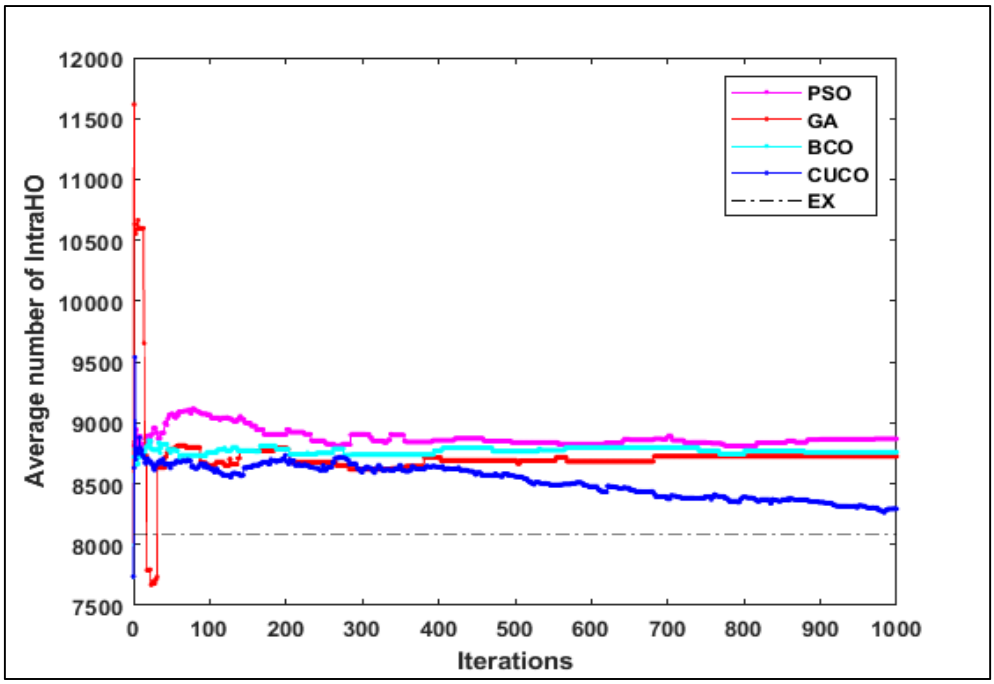

Figure 6.17 Average number of intra-handovers for 61 RRH

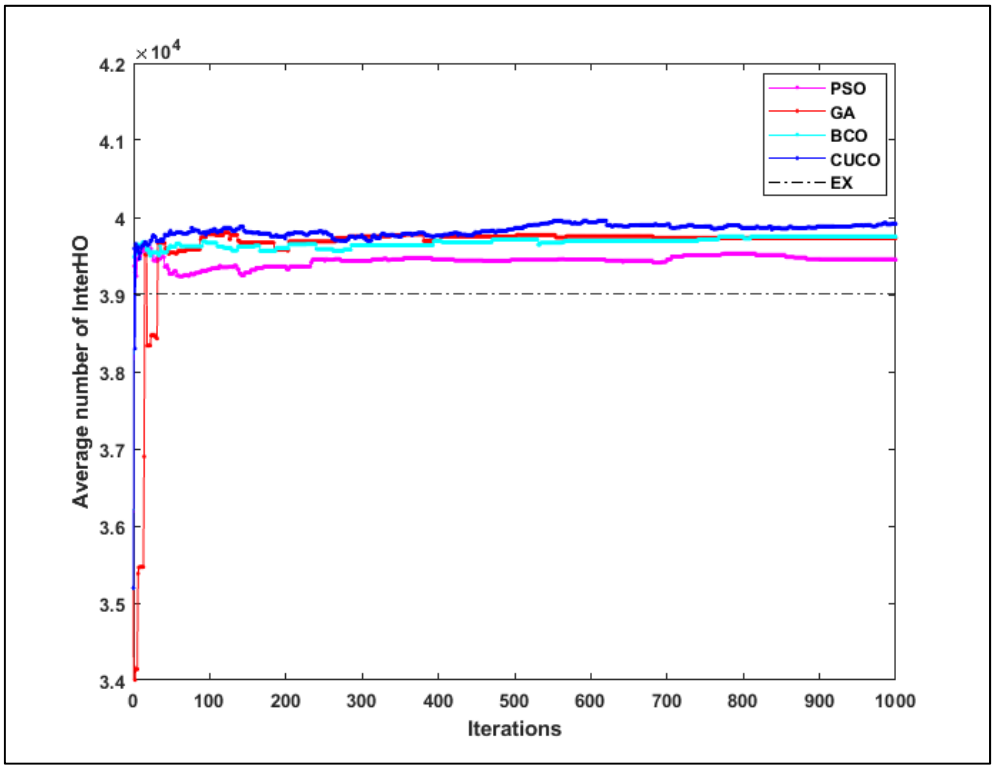

Figure 6.18 Average number of inter-handovers for 61 RRH

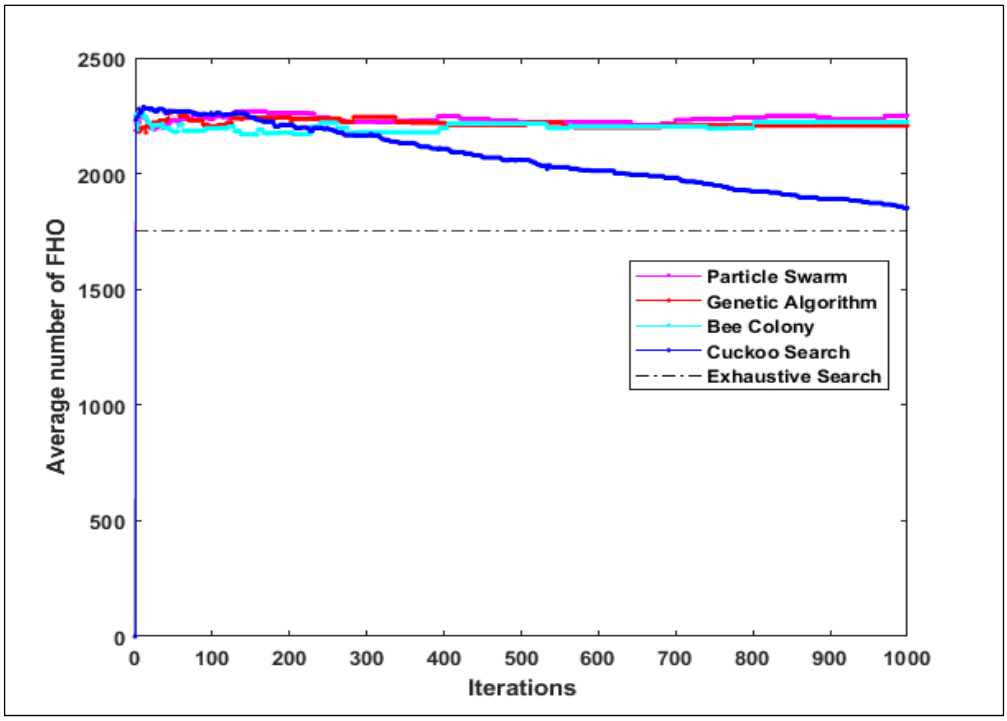

Figure 6.19 Average number of forced handovers for 61 RRH

To give a better clarification regarding the solution provided by BCO, PSO, CUCO, and GA algorithms. Table 6.9 indicates the solution from the best run obtained by the ES when compared to the evolutionary algorithms. It is clear from Figure 6.21 Sectors load that the CUCO succeeds in obtaining the optimum solution similar to the ES while the GA, BCO and PSO reach a very close solution to the optimum. The presented GA and PSO results that are different from the ES solution, correspond to one of the two runs, out of the 20 runs, that failed to provide the optimum solution. However, they still get almost 97% of the optimum solution provided by ES. Moreover, Table 6.9 shows the exact values obtained from this run. The CUCO maximizes the NP in the network by obtaining the optimum RRH-Sector-BBU allocation while GA and PSO succeed in reaching near optimum value during this specific run. Furthermore, the convergence of the BCO, PSO, CUCO, and GA algorithms is compared to the ES. ES reached the optimum solution after searching the whole space which is equivalent to 9^{61} iterations. On the other hand, the PSO, GA, BCO, and CUCO require only 683, 97, 188, and 700 iterations to converge, respectively. Therefore, it is clear how fast and accurate the solutions provided by the evolutionary algorithms are.

Table 6.9 Computational results for the best run out of the 20 runs for 61 RRH

KPIs vs algo.	NP Function value	Standard deviation	Convergence (iterations)	Number of active BBUs	BC
Initial	0.1045	87.3689	NA	4	210
BCO	0.7184	4.4381	188	4	0
GA	0.7142	20.3753	97	4	0
PSO	0.7207	10.0755	683	4	0
CUCO	0.7389	2.4618	700	4	0
Exhaustive	0.7389	2.4618	9^{61}	4	0

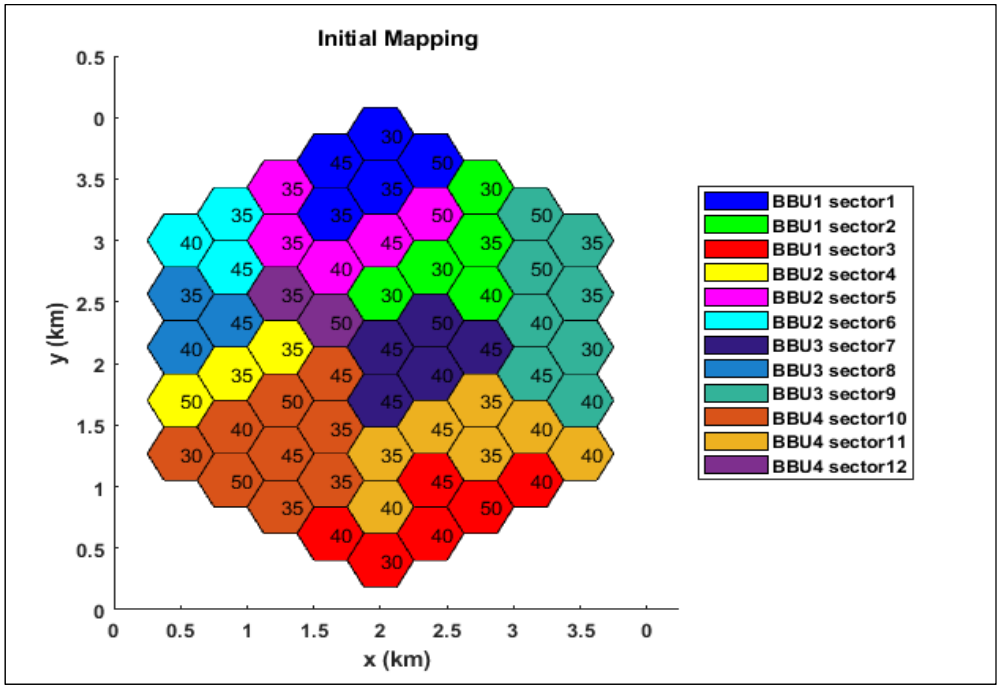

Figure 6.20 Initial map topology for 61 RRH at time t

Figure 6.21 shows that the load is fairly shared among the active sectors, which also reflects the minimization of the standard deviation by the CUCO, BCO, GA, and PSO to become 2.4618, 4.4381, 20.3753, and 10.0755 respectively, instead of 87.3689 obtained from the initial configuration. Consequently, more UEs could be admitted in different sectors in the future. Furthermore, Figure 6.22 shows the BBU load after obtaining the optimal RRH-Sector-BBU allocation. The number of used BBU is still 4, that is because the network load is 83% and the operation of all the BBUs are necessary to serve the current users in the network.

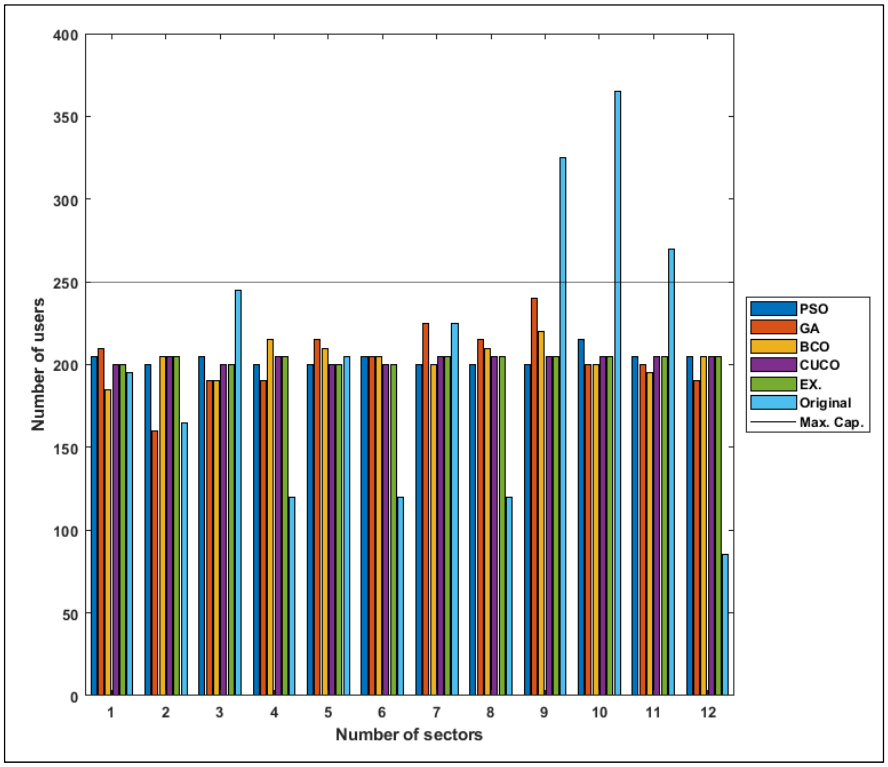

Figure 6.21 Sectors load for 61 RRH

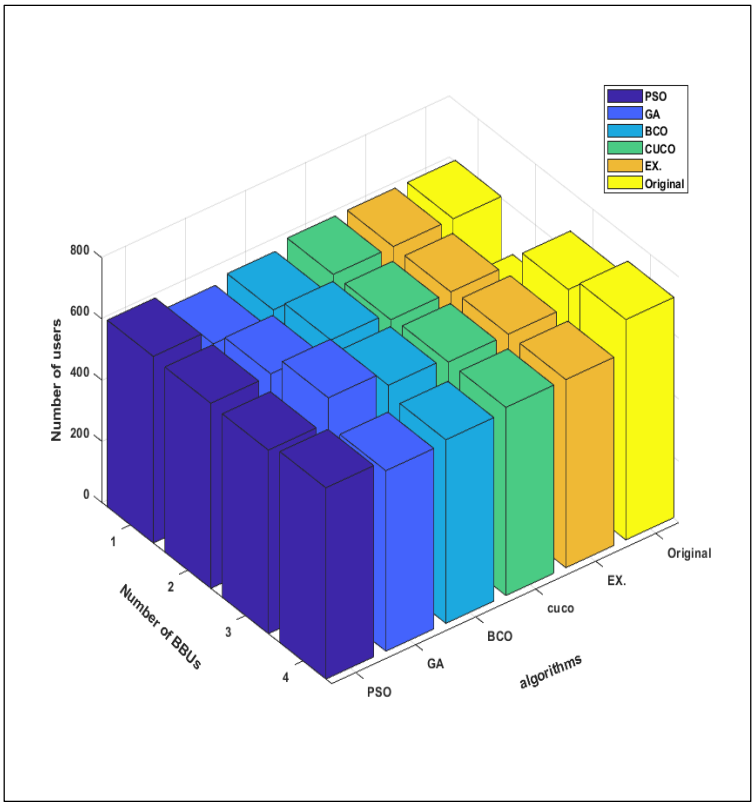

Figure 6.22 BBU load for 61 RRH

6.7 Chapter Summary

This chapter presents the simulation results for the RRH-Sector selection and load balancing framework using MATLAB software. The simulation results show that the MDP-P algorithm provides notably increased operator revenue when compared to commonly used network selection techniques based on the received signal strength (RSS). At the same time, the users blocking probabilities and the data rate are also notably improved.

Regarding the results for RRH-Sector-BBU mapping, it is observed that in the low load scenario the evolutionary algorithms succeed in decreasing the number of operating BBUs to two instead of three during the load balancing. Furthermore, for the low and high load scenarios, the optimum solution is reached 18 times out of 20 when the BCO, PSO, CUCO, and GA algorithms converge. Consequently, the NP is maximized, and a well-balanced network is achieved.

CONCLUSIONS AND FUTURE WORKS

Conclusions

In this thesis, C-RAN is introduced as a novel architecture that supports the tremendous increase in mobile network traffic. Despite the novel advancements that C-RAN offers, a time-varying traffic environment can cause load imbalances, resulting in inefficient resource utilization. Consequently, the network performance (NP) can degrade in terms of the blocked users, the number of unnecessary handovers, and the power consumption. Thus, the thesis objective was to present an RRH-Sector pair selection for new connections and network load-balancing framework that enhances the Quality of Service (QoS), the NP, and operator reward in C-RAN. In the first part of the framework, the RRH-Sector pair selection model provided several performance advantages for users and better utilization of the networks' resources. These benefits were achieved by developing the RRH-Sector pair selection model with the objective of maximizing the integrated operator-user utility. The decomposed Markov Decision Process (MDP) approach is used in the selection model. In particular, we use the sector shadow price, sector net gain, and RRH-sector net gain concepts to cope with complexity of the exact MDP model. The model integrates the objectives of the users and the networks' operators by combining their utilities. For users, we use a utility being a function of relevant QoS metrics. For the network operator, the utility is defined as the reward that can be also interpreted as the revenue. The integration of the user and operator utilities is achieved by introducing the operator reward penalty being a function of the user utility. In the second part of the framework, the load-balancing problem is addressed via optimization of the RRH-Sector-BBU dynamic mapping which formulated as a linear integer-based constrained optimization problem. The objective of this optimization is formulated using various NP KPIs such as the standard deviation of the sectors load, number of forced handovers blocked users, number of handovers, and the power consumption. And each KPI has a defined weight that could be chosen by the network operator according to its preferences. The solutions of the RRH-Sector-BBU dynamic mapping were obtained by using several evolutionary algorithms: BCO, CUCO, GA, and PSO.

The simulation results showed that the MDP based algorithm provided notably increased operator reward compared to commonly used network selection techniques based on the RSS. At the same time, the users blocking probabilities and the data rate are also notably improved. Regarding the dynamic RRH-Sector-BBU mapping optimization the numerical results showed that in the considered scenarios all evolutionary algorithms reached the optimum solution 18 times out of 20 cases. Moreover, the results in the selected scenario indicate that both CUCO and BCO succeed to obtain the optimum NP function value while the PSO and GA succeed in reaching near optimum value. Furthermore, the BCO is the fastest one to converge and reach the optimum solution, while the CUCO search took more time to converge to the optimum solution. On the other hand, both PSO and GA converged quickly, but they did not reach the optimum solution. Finally, the results show that the network performance is quite sensitive to the KPIs weights so their choice should be considered carefully by the network operator according to its preferences.

Future Work

Several possible future research directions are listed below:

- More power saving could be considered if the underutilized RRHs turned off; however, switching On/Off RRHs needs the cooperation from all RRHs in the network to guarantee UEs the required QoS and connectivity.
- The addition of the users' mobility to the proposed framework is going to make the model more practical, however, it will increase the complexity.
- Considering different types of schedulers to check how the resources are distributed differently and evaluate the performance of the framework in C-RAN.
- Finally, a comparison between the MDP based approach with an approach based on deep reinforcement learning (DRL). While both approaches (MDP and DRL) belong to the same dynamic programming family, the DRL based approach requires less information about the system dynamics. Still, its performance and adaptiveness to variable traffic conditions need to be verified.

APPENDIX

LIST OF PUBLICATIONS

Journals:

Submitted

Mouawad, M., Dziong, Z., & Mah F. (2021). RRH-Sector selection and load balancing based on MDP and Dynamic RRH-Sector-BBU Mapping in C-RAN. *IEEE Transactions on Network and Service Management*. IEEE.

Conferences:

Accepted and Published

Mouawad, M., Dziong, Z., & Addali, K. (2019). RRH selection and load balancing through Dynamic BBU-RRH Mapping in C-RAN. Dans *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)* (pp. 1-5). IEEE.

Mouawad, M., Dziong, Z., & El-Ashmawy, A. (2018). Load balancing in 5G C-RAN based on dynamic BBU-RRH mapping supporting IoT communications. Dans *2018 IEEE Global Conference on Internet of Things (GCIoT)* (pp. 1-6). IEEE.

Mouawad, M., Dziong, Z., & Khan, M. (2018). Quality of service aware dynamic BBU-RRH mapping based on load prediction using markov model in C-RAN. Dans *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (pp. 1907-1912). IEEE.

LIST OF BIBLIOGRAPHICAL REFERENCES

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

Ahmed, F., Dowhuszko, A. A., & Tirkkonen, O. (2016). Network optimization methods for self-organization of future cellular networks: Models and algorithms. Dans *Self-Organized Mobile Communication Technologies and Techniques for Network Optimization* (pp.s 35-65). IGI Global.

Alhumaima, R. S., Khan, M., & Al-Raweshidy, H. S. (2016). Component and parameterised power model for cloud radio access network. *IET Communications, 10*(7), 745-752. doi: 10.1049/iet-com.2015.0752

Alliance, N. (2013). Suggestions on potential solutions to C-RAN. *White Paper, January*.

Arslan, M. Y., Sundaresan, K., & Rangarajan, S. (2015). Software-defined networking in cellular radio access networks: potential and challenges. *IEEE Communications Magazine, 53*(1), 150-156.

Benaatou, W., Latif, A., & Pla, V. (2017). Vertical handover decision algorithm in heterogeneous wireless networks. *International Journal of Internet Protocol Technology (Online), 10*(4), 197-213.

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

Checko, A., Christiansen, H. L., Yan, Y., Scolari, L., Kardaras, G., Berger, M. S., & Dittmann, L. (2015). Cloud RAN for mobile networks—A technology overview. *IEEE Communications surveys & tutorials, 17*(1), 405-426.

Chen, L., Yang, D., Nogueira, M., Wang, C., & Zhang, D. (2020). Data-Driven C-RAN Optimization Exploiting Traffic and Mobility Dynamics of Mobile Users. *IEEE Transactions on Mobile Computing*.

Chen, X., Li, N., Wang, J., Xing, C., Sun, L., & Lei, M. (2014). A Dynamic Clustering Algorithm Design for C-RAN Based on Multi-Objective Optimization Theory. Dans *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)* (pp. 1-5). doi: 10.1109/VTCSpring.2014.7022775

Chen, Y.-S., Chiang, W.-L., & Shih, M.-C. (2018). A dynamic BBU–RRH mapping scheme using borrow-and-lend approach in cloud radio access networks. *IEEE Systems Journal, 12*(2), 1632-1643.

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

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

Cisco. (2016-2020, March 28, 2017). Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update. Repéré

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

CL I, C., Rowell, S. H., & Z Xu, G. L. (2014). Z. Pan,"Toward green and soft: A 5G perspective,". *IEEE Communications Magazine, 52*(2), 66-73.

Cunhua, P., Huiling, Z., Gomes, N. J., & Jiangzhou, W. (2017). Joint Precoding and RRH Selection for User-Centric Green MIMO C-RAN. *IEEE Transactions on Wireless Communications, 16*(5), 2891-2906. doi: 10.1109/TWC.2017.2671358. Repéré à http://dx.doi.org/10.1109/TWC.2017.2671358

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

Dhifallah, O., Dahrouj, H., Al-Naffouri, T. Y., & Alouini, M. (2015). Joint Hybrid Backhaul and Access Links Design in Cloud-Radio Access Networks. Dans *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)* (pp. 1-5).

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

Dziong, Z., Choquette, J., Liao, K.-Q., & Mason, L. (1990). Admission control and routing in ATM networks. *Computer Networks and ISDN Systems, 20*(1-5), 189-196.

El Fachtali, I., Saadane, R., & ElKoutbi, M. (2016). Vertical handover decision algorithm using ants' colonies for 4G heterogeneous wireless networks. *Journal of Computer Networks and Communications, 2016*.

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

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

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

Ge, X., Ye, J., Yang, Y., & Li, Q. (2016). User mobility evaluation for 5G small cell networks based on individual mobility model. *IEEE Journal on Selected Areas in Communications, 34*(3), 528-541.

Grassmann, W. K. (1977). Transient solutions in Markovian queueing systems. *Computers & Operations Research, 4*(1), 47-53.

Han, B., Liu, L., Zhang, J., Tao, C., Qiu, C., Zhou, T., . . . Piao, Z. (2019). Research on Resource Migration Based on Novel RRH-BBU Mapping in Cloud Radio Access Network for HSR Scenarios. *IEEE Access, 7*, 108542-108550.

Hashim, W., Ismail, A., Abd Ghafar, N., & Dzulkifly, S. (2013). Cognitive Selection Mechanism Performance in IEEE 802.11 WLAN. *International Journal of Computer and Communication Engineering, 2*(4), 477.

He, W., Gong, J., Su, X., Zeng, J., Xu, X., & Xiao, L. (2016). SDN-enabled C-RAN? An intelligent radio access network architecture. Dans *New Advances in Information Systems and Technologies* (pp. 311-316). Springer.

Holma, H., & Toskala, A. (2009). *LTE for UMTS: OFDMA and SC-FDMA based radio access*. John Wiley & Sons.

Hoydis, J., Ten Brink, S., & Debbah, M. (2011). Massive MIMO: How many antennas do we need? Dans *2011 49th Annual Allerton conference on communication, control, and computing (Allerton)* (pp. 545-550). IEEE.

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

Hwang, I., Song, B., & Soliman, S. S. (2013). A holistic view on hyper-dense heterogeneous and small cell networks. *IEEE Communications Magazine, 51*(6), 20-27.

Kaiwei, W., Wuyang, Z., & Shiwen, M. (2017). On Joint BBU/RRH Resource Allocation in Heterogeneous Cloud-RANs. *IEEE Internet of Things Journal, 4*(3), 749-759. doi: 10.1109/JIOT.2017.2665550. Repéré à http://dx.doi.org/10.1109/JIOT.2017.2665550

Khan, M., Alhumaima, R. S., & Al-Raweshidy, H. S. (2017). QoS-Aware Dynamic RRH Allocation in a Self-Optimised Cloud Radio Access Network with RRH Proximity Constraint. *IEEE Transactions on Network and Service Management, PP*(99), 1-1. doi: 10.1109/TNSM.2017.2719399

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

Lin, K., Wang, W., Zhang, Y., & Peng, L. (2017). Green spectrum assignment in secure cloud radio network with cluster formation. *IEEE Transactions on Sustainable Computing*.

Lin, Y., Shao, L., Zhu, Z., Wang, Q., & Sabhikhi, R. K. (2010). Wireless network cloud: Architecture and system requirements. *IBM Journal of Research and Development, 54*(1), 4: 1-4: 12.

Liu, L., & Yu, W. (2017). Cross-layer design for downlink multihop cloud radio access networks with network coding. *IEEE Transactions on Signal Processing, 65*(7), 1728-1740.

Liu, Q., Wu, G., Guo, Y., Zhang, Y., & Hu, S. (2016). Energy Efficient Resource Allocation for Control Data Separated Heterogeneous-CRAN. Dans *2016 IEEE Global Communications Conference (GLOBECOM)* (pp. 1-6).

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

Love, R., Kuchibhotla, R., Ghosh, A., Ratasuk, R., Xiao, W., Classon, B., & Blankenship, Y. (2008). Downlink control channel design for 3GPP LTE. Dans *2008 IEEE Wireless Communications and Networking Conference* (pp. 813-818). IEEE.

LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures. 3GPP

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

Mahardhika, G., Ismail, M., & Nordin, R. (2015). Vertical handover decision algorithm using multicriteria metrics in heterogeneous wireless network. *Journal of Computer Networks and Communications, 2015*.

Mishra, D., Amogh, P. C., Ramamurthy, A., Franklin, A. A., & Tamma, B. R. (2016). Load-aware dynamic RRH assignment in Cloud Radio Access Networks. Dans *2016 IEEE Wireless Communications and Networking Conference* (pp. 1-6). doi: 10.1109/WCNC.2016.7564824

Miyim, A., Ismail, M., Nordin, R., & Mahardhika, G. (2013). Generic vertical handover prediction algorithm for 4G wireless networks. Dans *2013 IEEE International Conference on Space Science and Communication (IconSpace)* (pp. 307-312). IEEE.

Mobile, C. (2011). C-RAN: the road towards green RAN. *White paper, ver, 2*, 1-10.

Namba, S., Warabino, T., & Kaneko, S. (2012). BBU-RRH switching schemes for centralized RAN. Dans *7th International Conference on Communications and Networking in China* (pp. 762-766). IEEE.

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

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

Ngah, S., & Bakar, R. A. (2017). Sigmoid function implementation using the unequal segmentation of differential lookup table and second order nonlinear function. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9*(2-8), 103-108.

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

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

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

Pan, C., Zhu, H., Gomes, N. J., & Wang, J. (2017). Joint precoding and RRH selection for user-centric green MIMO C-RAN. *IEEE Transactions on Wireless Communications, 16*(5), 2891-2906.

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

Peng, M., Sun, Y., Li, X., Mao, Z., & Wang, C. (2016). Recent advances in cloud radio access networks: System architectures, key techniques, and open issues. *IEEE Communications surveys & tutorials, 18*(3), 2282-2308.

Peng, M., Zhang, K., Jiang, J., Wang, J., & Wang, W. (2015). Energy-Efficient Resource Assignment and Power Allocation in Heterogeneous Cloud Radio Access Networks. *IEEE Transactionson Vehicular Technology, 64*(11), 52755287.

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

Quek, T. Q., & Yu, W. (2017). *Cloud radio access networks: Principles, technologies, and applications*. Cambridge University Press.

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

Schweitzer, P. J., & Federgruen, A. (1979). Geometric convergence of value-iteration in multichain Markov decision problems. *Advances in Applied Probability, 11*(1), 188-217.

Sevcik, P. (2002). Understanding how users view application performance. *Business Communications Review, 32*(7), 8-9.

Shaw, J. A. (2013). Radiometry and the Friis transmission equation. *American journal of physics, 81*(1), 33-37.

Shi, Y., Zhang, J., & Letaief, K. B. (2015). Robust Group Sparse Beamforming for Multicast Green Cloud-RAN With Imperfect CSI. *IEEE Transactions on Signal Processing, 63*(17), 4647-4659. doi: 10.1109/TSP.2015.2442957

Simeone, O., Maeder, A., Peng, M., Sahin, O., & Yu, W. (2016). Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems. *Journal of Communications and Networks, 18*(2), 135-149.

Sriharsha, M., Dama, S., & Kuchi, K. (2017). A complete cell search and synchronization in LTE. *EURASIP Journal on Wireless Communications and Networking, 2017*(1), 101.

Srinivas, M., & Patnaik, L. M. (1994). Genetic algorithms: A survey. *computer, 27*(6), 17-26.

Sundaresan, K., Arslan, M. Y., Singh, S., Rangarajan, S., & Krishnamurthy, S. V. (2016). FluidNet: A flexible cloud-based radio access network for small cells. *IEEE/ACM Transactions on Networking (TON), 24*(2), 915-928.

Tohidi, M., Bakhshi, H., & Parsaeefard, S. (2020). Flexible Function Splitting and Resource Allocation in C-RAN for Delay Critical Applications. *IEEE Access*.

Tsagkaris, K., Poulios, G., Demestichas, P., Tall, A., Altman, Z., & Destré, C. (2015). An open framework for programmable, self-managed radio access networks. *IEEE Communications Magazine, 53*(7), 154-161.

Tschofenig, H., Hodges, J., & Peterson, J. (2010). Nokia Siemens Networks.

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

Venkataraman, H., & Trestian, R. (2017). *5G Radio Access Networks: centralized RAN, cloud-RAN and virtualization of small cells*. CRC Press.

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

Wang, K., Zhou, W., & Mao, S. (2017). On joint BBU/RRH resource allocation in heterogeneous cloud-RANs. *IEEE Internet of Things Journal, 4*(3), 749-759.

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

Yang, C., Chen, Z., Xia, B., & Wang, J. (2015). When ICN meets C-RAN for HetNets: an SDN approach. *IEEE Communications Magazine, 53*(11), 118-125.

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

You, L., & Yuan, D. (2019). User-centric Performance Optimization with Remote Radio Head Cooperation in C-RAN. *IEEE Transactions on Wireless Communications*.

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

Yu, Z., Wang, K., Ji, H., Li, X., & Zhang, H. (2016). Dynamic resource allocation in TDD-based heterogeneous cloud radio access networks. *China Communications, 13*(6), 1-11.

Yuh-Shyan, C., Wen-Lin, C., & Min-Chun, S. (2018). A Dynamic BBU-RRH Mapping Scheme Using Borrow-and-Lend Approach in Cloud Radio Access Networks. *IEEE Systems Journal, 12*(2), 1632-1643. doi: 10.1109/JSYST.2017.2666539. Repéré à http://dx.doi.org/10.1109/JSYST.2017.2666539

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

Zhang, Y., Wang, S., & Ji, G. (2015). A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems in Engineering, 2015*.

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

Zhao, M.-M., Cai, Y., Zhao, M.-J., & Champagne, B. (2019). Joint Content Placement, RRH Clustering and Beamforming for Cache-Enabled Cloud-RAN. Dans *ICC 2019-2019 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.

Zhou, S., Zhao, M., Xu, X., Wang, J., & Yao, Y. (2003). Distributed wireless communication system: a new architecture for future public wireless access. *IEEE Communications Magazine, 41*(3), 108-113.

Zhu, F., & McNair, J. (2006). Multiservice vertical handoff decision algorithms. *EURASIP Journal on wireless communications and networking, 2006*(1), 025861.